

Empirical Testing of the Capital Asset Pricing Model (CAPM): An Assessment of Its Predictive Ability in Contemporary Financial Markets

Rongsen Chen

Business School, University of Birmingham, Birmingham, UK

Email: sen_0725@qq.com

How to cite this paper: Chen, R. S. (2026). Empirical Testing of the Capital Asset Pricing Model (CAPM): An Assessment of Its Predictive Ability in Contemporary Financial Markets. *Modern Economy*, 17, 1297-1323.

<https://doi.org/10.4236/me.2026.1710063>

Received: June 22, 2026

Accepted: October 16, 2026

Published: October 19, 2026

Copyright © 2026 by author(s) and

Scientific Research Publishing Inc.

This work is licensed under the Creative

Commons Attribution International

License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This dissertation analyses the empirical forecasting power of CAPM for seven developed and emerging markets from 2003 to 2024. While the model has been a major contributor to finance theory and practice, over the last few decades, evidence has emerged that suggests it is not a good representation of the risk-return relationship in today's financial markets. The study creates beta sorted portfolios using monthly returns of large listed stocks, tests the CAPM against the three factor and five factor models proposed by Fama and French, conducts a time series test, a Fama and MacBeth cross sectional regression, a rolling out of sample forecasting exercise, and analyzes two periods of crisis. What the results show is a definite conflict. The strict CAPM is rejected statistically in most markets, as the pricing errors are picked up by the multifactor models as they should be. The ordering, by the way, is swapped out of sample and the single factor CAPM generates the least out of sample prediction errors in 4 of the 7 markets. In times of crises, the level of predictability falls, mostly due to varying betas in less integrated markets. The results indicate that statistical rejection and practical usefulness are separable, and that the CAPM is a valid forecasting instrument.

Keywords

CAPM Model, Prediction, Accuracy

1. Introduction

1.1. Research Background

One of the more famous models in financial economics is the Capital Asset Pricing Model (CAPM). The model is based on modern portfolio theory, which suggests that the expected return of any risk asset is related to the asset's sensitivity

to non-diversifiable market risk. Beta indicates how sensitive an asset's return is to changes in the return of the market portfolio, and it represents this exposure. The CAPM pays attention to the systematic risk of an investment only because the asset-specific risk can be reduced by diversifying as described (Sharpe, 1964; Lintner, 1965). The resulting linear relationship between beta and expected return makes it easy to have a clear theoretical basis to estimate required returns and cost of equity.

A key attraction of the CAPM is its simplicity. The complex valuation problem is broken down into three observable or estimable components: the risk-free rate, and an asset's beta and the expected market risk premium. This has led to the continued use of the model in corporate valuation, capital budgeting, performance measurement and regulatory decision making. However, it has been doubted that it is an empirical fact for decades. Assumptions needed to obtain the model are frictionless trading, unrestricted borrowing and lending at a common risk-free interest rate, homogeneous expectations and the existence of a mean-variance-efficient market portfolio. These conditions are likely to be violated in real financial markets, and this makes it questionable whether the theoretical model can be used to realistically assess realised asset returns (Roll, 1977; Fama & French, 2004).

Empirical evidence also has shown patterns in average stock returns that are not adequately explained by the market beta alone. The initial tests typically demonstrated that there is a positive relationship between risk and return, but the Security Market Line was generally less steep than the estimated CAPM. In practice, low beta stocks had higher returns and high beta stocks lower returns than the model predicted (Black, Jensen, & Scholes, 1972; Fama & MacBeth, 1973). Subsequent studies uncovered further return patterns linked to the book-to-market equity, profitability, investment and momentum of the firm. These findings paved the way to the construction of multi-factor models and doubts were cast on the idea that beta was the only factor affecting expected returns (Fama & French, 1992; Carhart, 1997; Fama & French, 2015).

Much has happened in the world of financial markets since the CAPM was developed. A new era of electronic trading, fast information flow and increasingly global capital flows, combined with institutional and algorithmic investors, has transformed market dynamics. The sharp volatility and changes in correlations and risk premiums during times of stress were further illustrated during the global financial crisis of 2007 - 2009 and during the disruption of the markets caused by the coronavirus (COVID-19) pandemic. In light of these developments, the question of the relevance of a static single-factor model under different market and economic environments should be revisited, not only within the country but also across countries.

The CAPM is not a thing of the past in academic instruction and practice, however, the criticisms it has received have not gone away. Its long standing use is driven both by its economic simplicity of use and by the minimal amount of data required. Multi-factor models can be better at explaining the data in the sample,

but they are also more complex in terms of the number of factors that need to be chosen and estimated and some factors may not have an obvious equilibrium interpretation. In this sense, the CAPM is still a valuable standard to check the accuracy of more complex models. An empirical reassessment should not just consider whether the CAPM is entirely accurate, but rather when it is reasonably accurate and when it is not and whether modifying the CAPM with other functional forms produces meaningful improvements that are economically valuable.

The aim of this research is to test the predictive validity of the CAPM with a geographically spread sample of listed stocks from developed and emerging markets from January 2000 to December 2024. The analysis will be carried out over the entire sample and major periods of financial stress. The results of CAPM will also be benchmarked against some of the multi-factor models which are used, specifically the Fama-French 3 factor model and the Fama-French 5 factor model. Using both cross-market evidence and analysis of the crisis period, the study aims to contribute to our appraisal of the CAPM relevance to the new financial theory and practice.

1.2. Research Objectives

The main goal is to empirically test the power of the CAPM to explain and predict equity returns under various financial market and economic environments. Instead of repeating an individual market test, the study is designed like a comparative test, which includes a number of markets, an alternative asset-pricing model and market regimes.

- To estimate the relationship between systematic risk, measured by beta, and subsequent equity returns across developed and emerging markets.
- To check if the estimated intercept and market-risk-premium are consistent with the theoretical constraints in the CAPM.
- To compare CAPM with the Fama-French three factor model and five factor model in terms of their explanatory and predictive ability.
- To study if performance of CAPM is different during crisis periods such as the global financial crisis and the COVID-19 market shock.
- To evaluate if the empirical performance of the model is different across markets due to variation in market development, liquidity and institutional conditions.
- To determine when the CAPM is still a viable practical model even in the presence of more sophisticated models.

2. Literature Review

The CAPM theory was based on Markowitz's (1952) mean-variance theory. In the example, Markowitz showed that investors should consider the expected return of a portfolio in addition to variance and covariance of returns. Diversification can help investors remove a significant amount of the risk associated with the firm they invest in, creating an efficient frontier of portfolios of highest expected return

at a given level of total risk. The framework was later expanded by Sharpe (1964) and Lintner (1965) to a market-equilibrium model. All investors are investing in efficient portfolios and they have identical expectations and assumptions from the model, they pool the risk free asset with the market portfolio. The return of each risky asset is then adjusted for its contribution to the volatility of this market portfolio, and not for its own volatility.

According to the standard CAPM, beta represents the ratio of the covariance of a particular asset's return to the market return and the variance of the market return. The model assigns that there is a positive linear association between beta and the expected excess return, the Security Market Line. An asset that has a beta greater than one should have a higher expected return than the market as it has a greater volatility than the market, while an asset with a beta less than one should have a lower expected return. The model also suggests that the regression intercept should be zero if returns are returns in excess of the risk free rate, and that none of the other firm characteristics should be explained by the expected returns after accounting for beta.

The predictions were mixed early on. Black, Jensen and Scholes (1972) discovered evidence of a positive relation between beta and average return, but the estimated relation was flatter than the Sharpe-Lintner CAPM would have predicted. Fama and MacBeth (1973) designed two-stage cross-sectional regression and also observed that portfolios with higher betas were more likely to have higher returns. Their results, however, did not completely meet the limitations of the model. Equities with lower betas had generally higher returns and equities with higher betas had generally lower returns than the Security Market Line had forecasted. These results implied that beta carried return information, but was not a comprehensive representation of the cross section.

One of the difficulties in testing the CAPM is that the actual market portfolio is unknown. It should contain all the assets that investors have that pose a risk, such as stocks, bonds, real estate, private businesses, and human capital. Empirical tests compare instead with indices of the stock market. It may therefore be concluded that a rejection of the CAPM may be due to using an inefficient market proxy and not necessarily to the failure of the model. On the other hand, a successful test can result in the determination that the chosen index is "more-or-less" mean-variance efficient. This critique relates to the fact that the CAPM cannot be tested in isolation with respect to the assumptions made about the market portfolio being analyzed.

2.1. Objections

The most compelling empirical objection came from the work on return anomalies. Fama and French (1992) found that size and the book-to-market ratio accounted for significant cross-sectional variation in average stock returns, with little additional explanatory power added by the market beta. Later, Fama and French introduced the size factor and the value factor in their three factor model,

and it greatly improved the ability to explain diversified portfolio returns (Fama & French, 1993). Carhart (1997) added a momentum factor, an indication of how well a stock has performed in the recent past, which is likely to retain these attributes over intermediate time horizons. Fama and French (2015) later extended their model to incorporate profitability and investment factors.

If the MFI is better fit, it does not necessarily imply that the extra factors are fundamental risk factors. Investor behaviour, friction or patterns identified after many observations of past data could be some of the factors captured. But, Harvey, Liu and Zhu (2016) caution that a large search of factors can lead to results that are statistically significant but economically fragile. When contrasting the CAPM with richer models, this concern is very relevant as it is likely that the addition of variables will lead to higher in-sample explanatory power, but that improvement should also be clear out of sample and market and period.

The other is the difference between expected and realised returns. Empirical studies typically rely on realised returns as noisy proxies for expected returns, while asset-pricing theories refer to expected returns. Under the assumption that the information shocks are unexpected, there is possible a difference between realised and expected average returns as argued by Elton (1999). The estimated risk-return relationship can be weakened by measurement error in beta and in realised returns. These problems have been reduced by the processes of portfolio formation and longer estimation periods, as well as by the Fama-MacBeth procedure, but not entirely eliminated.

Conditional versions of the CAPM make an attempt to relate to the fact that betas and market risk premiums can change over time. Investor risk aversion and firms' operating leverage can shift, as do economic conditions, meaning that a beta calculated from past observations may be different than the one of interest for future situations. Even if there is plausible variation in conditional betas, however, there are several important pricing anomalies that are not explained by Lewellen and Nagel (2006). This suggests that, in some contexts, time variation may be beneficial to the model, but does not necessarily recover its empirical validity.

There are further challenges in the international and emerging-market evidence. Market integration could be incomplete, and returns could be impacted by currency exposure, sovereign risk, political instability, ill liquidity and restrictions on foreign investment. A domestic share index may therefore not be a good proxy of the relevant market stock index and vice versa, a global share index might lack local segmentation. Such features give a compelling argument to run the model on markets, not generalise from evidence from the U.S. only.

2.2. Factors Affecting CAPM Predictability in Contemporary Financial Markets

The reason why the CAPM is found to predict returns well in some markets and poorly in others is not due to any single shortcoming of the CAPM but rather the environment in which it is called upon to perform. It is violated by several aspects

of contemporary markets and the errors in pricing are hard to identify as originating from one source.

The simplest problem is that the parameters which the model considers to be fixed, are not. A beta may be very different from one year to the next, as there are periods when the business cycle swings in the opposite direction, with changes in leverage and asset mix of firms and changes in investor risk aversion. [Lewellen and Nagel \(2006\)](#) demonstrate that only at the margins is allowing for realistic time variation useful: the changes in betas and premiums are, simply, not large enough to explain the bulk of the model errors in pricing. When times get tough, the problem only gets sharper. [Longin and Solnik \(2001\)](#) report that correlations between international equity markets increase sharply in periods of bear markets, meaning that betas calculated during periods of calm may underestimate exposures during periods that are most important for accurate risk measurement, as in the cases studied here.

The investor behaviour exacerbates the estimation problem. But those investors that cannot borrow freely reach for return by tilting their portfolios towards high-beta stocks, driving up the price of these stocks and flattening the Security Market Line, so that their betting-against-beta portfolios outperform the Sharpe-Lintner prediction by significant positive amounts in equity markets around the world, says [Frazzini and Pedersen \(2014\)](#). By a different path, [Baker, Bradley and Wurgler \(2011\)](#) reach a similar conclusion: they argue that institutional pressures discourage variation from the norm and that there is a behavioral preference for lottery-like payments. A still more intriguing conundrum is added by [Ang et al. \(2006\)](#): stocks with high idiosyncratic volatility post abnormally low returns, a pattern that cannot possibly be explained by the pricing of systematic risk.

Frictions matter too. [Amihud \(2002\)](#) demonstrates that expected returns increase as illiquidity increases and that “market-wide” liquidity shocks shift returns over time, so that some of the intercept from a single factor regression is likely just compensation for holding assets that are not easily tradable. For emerging markets, where trading is light and access can be limited, [Bekaert, Harvey and Lundblad \(2007\)](#) find that local liquidity forecasts returns even after controlling for traditional risk factors.

The trading environment itself is also a different one since the 60's. Today, much of the order flow is done so through algorithmic and high-frequency traders, which has also impacted the short-horizon return dynamics on which betas are estimated, as noted by [Hendershott, Jones and Menkveld \(2011\)](#) that this affects the informativeness of quotes and has reduced liquidity. Passive investing is the opposite: [Ben-David, Franzoni and Moussawi \(2018\)](#) illustrate that shares that are in the hands of a large number of exchange-traded funds (ETFs) tend to be more volatile and more sensitive to fund-level demand shocks, thus gaining in comovement with other ETF shares but not necessarily in fundamental systematic risk.

The question of which market portfolio is the right one is also raised. As [Bekaert](#)

and Harvey (1995) describe, the extent to which emerging markets are integrated with world capital markets varies with time and returns in emerging markets are a mixture of local and global risk. A domestic CAPM in an integrated market or a world CAPM in a segmented market misprices the risk and fails to predict.

But it doesn't imply that all this failure in the model is purely economic. The true market portfolio is unobserved (Roll, 1977), realised returns are noisy proxies for expectations (Elton, 1999), and Harvey, Liu and Zhu (2016) warn that the apparent outperformance of a lot of rival factors is largely due to data mining. But, predictability is the product of instability of the parameters, limited investors, illiquidity, altered market structure, incomplete integration, and imperfect measurement, in short. The design of this study, across markets, regimes and competing models, will seek to uncover which of these forces is the strongest in practice.

2.3. The Continuing Relevance of the CAPM

The large number of problems with the strict CAPM does not mean that the model has no practical use. Firstly, it offers a clear connection between systematic risk and required return to diversified investors. Even though practitioners make adjustments to the estimated beta, risk-free rate or the risk-free rate differential, this logic continues to form the core of corporate finance. Communication and challenge of the inputs of the model are easier than with highly parameterised alternatives it can be said. Additionally, according to Lewellen and Nagel (2006), the CAPM is a theoretical and empirical reference. The usefulness of the performance of a multi-factor model is partly due to its ability to be compared to the simpler market model. Transparency, consistency and implementation cost may make the CAPM more suited to various applications if there is limited out-of-sample gain. For instance, Bartholdy and Pearce (2005) conclude that the usefulness of applying the Fama-French model to estimate the expected returns of individual firms might be a little small compared to the extra estimation costs involved.

According to Harvey et al. (2016), some "failures" could occur due to the practical problems of empiricism. It is not possible to observe the market portfolio completely, estimates of beta are noisy, realised returns are noisy estimates of expectations and the risk premium applicable to the portfolio could vary over time. Such restrictions should not be taken as evidence that the CAPM is right, but as a warning to avoid demanding every empirical failure to confirm the economic intuition that is at the heart of the CAPM. This is the reason that, in this paper, the CAPM is not interpreted as a theory to be taken for granted nor is it regarded as a model that is no longer relevant. It will test the model with the regular data and also through the use of portfolio level tests and alternative market proxies as well as comparing the model to the existing factor models. A series of tests during the crisis period will be performed to evaluate whether predictive errors also increase during significant changes in correlation and risk premiums. The study will also discriminate between statistical and economic improvements, and thus offer a balanced view of the usefulness of the CAPM in today's financial markets.

3. Methodology

This chapter sets out how the predictive ability of the CAPM will be tested. The design is based on the research goals: the study does not aim to replicate a single-market test, but to design the study as a comparison among markets, models and market regimes, separating the conditions under which the CAPM is able to perform well from the conditions under which it does not.

3.1. Research Design

The approach of this study is quantitative, deductive. The CAPM makes clear and verifiable predictions: Excess returns are directly related to beta; the intercept of that relationship should be zero; and there should be no other characteristic besides beta that is receiving a premium. These predictions are considered testable hypotheses and the empirical work is aimed at determining not only that they are correct, but also at determining when and where. The analysis is based entirely on secondary data from the market, and the philosophy is positivist, i.e., to test theory against evidence, not the interpretation of market perceptions.

The analysis is organised in three comparisons. The model is estimated in developed and emerging markets separately, it is benchmarked against the Fama-French three factor model, five factor model, and it is re-estimated during specific time periods of market stress. The comparisons are related to the research goals presented in Chapter 1.

3.2. Data and Sample

The sample consists of monthly total returns for listed equities over the period January 2000 to December 2024, a span of 300 months that covers two major crises and several distinct market regimes. Four developed markets (the United States, the United Kingdom, Japan and Germany) and three emerging markets (India, Brazil and South Africa) are included; Chinese Mainland was dropped at the data-collection stage because return histories of sufficient length and quality could not be obtained. These markets were selected for their size, the availability of reliable return histories over the full period, and the variation they offer in liquidity, institutional quality and openness to foreign investment, which the literature review identified as important conditioning factors. The final universe comprises large, liquid constituents of each market's main index: thirty stocks in the United States, twenty each in the United Kingdom, Japan, Germany and India, fifteen in Brazil and fourteen in South Africa. Because the rolling beta-estimation window consumes the opening years of data, the first portfolios are formed at the end of 2002 and the effective test period runs from January 2003 to December 2024, giving 264 monthly observations per market.

Return data are drawn from Yahoo Finance, while factor returns and the one-month United States Treasury bill rate come from the Kenneth French data library, with the regional emerging-market factor series applied to the emerging

markets. Returns are computed in local currency, while the factor series are denominated in United States dollars; the implications of this mismatch are discussed in Section 3.5. A stock enters the estimation once it has at least twenty-four months of return data. Because the data source provides histories only for currently listed securities, delisted firms are absent from the sample, and the resulting survivorship bias is acknowledged as a limitation rather than corrected. The two crisis windows are fixed in advance rather than identified from the data: July 2007 to March 2009 for the global financial crisis, and February 2020 to December 2020 for the COVID-19 shock and its initial recovery. To keep the comparison currency-consistent and avoid mechanically disadvantaging the multifactor models, all returns, the market proxy and the risk-free rate are expressed in United States dollars to match the dollar-denominated Fama-French factor series before estimation (**Table 1**).

Table 1. Variable description and data sources.

Variable	Definition and construction	Source
Monthly stock return	Arithmetic monthly return computed from adjusted closing prices, which incorporate dividends and stock splits	Yahoo Finance
Market return	Monthly return on the broadest value-weighted national index (S&P 500, FTSE 100, Nikkei 225, DAX, BSE Sensex, Bovespa); for South Africa, the equal-weighted average return of the sampled stocks	Yahoo Finance
Risk-free rate (RF)	One-month United States Treasury bill rate	Kenneth French Data Library
Excess returns	Stock, portfolio and market returns minus the contemporaneous risk-free rate	Computed
Market factor (Mkt-RF)	Value-weighted regional market return in excess of the one-month T-bill rate, in US dollars	Kenneth French Data Library
Size factor (SMB)	Small-minus-big return spread from 2×3 size and book-to-market sorts (Fama & French, 1993)	Kenneth French Data Library
Value factor (HML)	High-minus-low book-to-market return spread from the same 2×3 sorts	Kenneth French Data Library
Profitability factor (RMW)	Robust-minus-weak operating profitability return spread (Fama & French, 2015)	Kenneth French Data Library
Investment factor (CMA)	Conservative-minus-aggressive asset-growth return spread (Fama & French, 2015)	Kenneth French Data Library
Pre-ranking beta	Slope from a rolling 60-month regression of stock excess returns on market excess returns, minimum 24 observations	Computed
Post-ranking beta	Slope from a full-sample regression of portfolio excess returns on market excess returns	Computed
Crisis indicator	Equal to one in July 2007 to March 2009 and February to December 2020, zero otherwise; windows fixed ex ante	Author-defined

3.3. Portfolio Formation and Beta Estimation

The errors-in-variables bias is well known for using betas directly in cross-sectional regressions and betas estimated for individual stocks are noisy. The study therefore adopts the Black, Jensen, & Scholes (1972) approach of portfolio. For

each stock, pre-ranking betas are estimated from the last sixty months of returns, and stocks in each market are placed into ten beta-ranked portfolios, rebalanced each year and post-ranking betas are estimated for each portfolio over the entire sample. In the primary analysis portfolios are value-weighted while in robustness checking they are equal-weighted. The broadest value weighted national indices available is used for the market portfolio (whereas the MSCI World index is used in integration tests). This treatment would not avoid the criticism of Roll (1977) that none of the observed proxies is complete, but it does show the sensitivity of the results to the proxy chosen.

3.4. Estimation and Hypothesis Testing

The empirical work rests on four regression specifications. For each beta-sorted portfolio i in month t , the CAPM is estimated as the time-series regression

$$R(i, t) - Rf(t) = a(i) + b(i)[Rm(t) - Rf(t)] + e(i, t),$$

Where

$R(i, t)$ is the portfolio return,

$Rf(t)$ the risk-free rate,

$Rm(t)$ the market return,

$b(i)$ the market beta

$a(i)$ the pricing error (alpha).

The Fama-French three-factor model adds the size and value factors, $R(i, t) - Rf(t) = a(i) + b(i)[Rm(t) - Rf(t)] + s(i)SMB(t) + h(i)HML(t) + e(i, t)$, and the five-factor model adds the profitability and investment factors $RMW(t)$ and $CMA(t)$ to the same equation. The cross-sectional relationship is estimated by the Fama-MacBeth second-pass regression $R(i, t) - Rf(t) = g0(t) + g1(t)b(i) + u(i, t)$, run separately each month, where $g0(t)$ is the intercept and $g1(t)$ the estimated price of beta risk. Table 2 reports the results of the time-series models (CAPM, three-factor and five-factor), Table 3 reports the cross-sectional Fama-MacBeth model, and Table 4 reports the out-of-sample prediction errors of all three models.

There are three explicit criteria to justify the models. The first criterion is cross-sectional consistency—the average intercept $g0$ must not differ significantly from zero, and the average slope $g1$ must be equal to the mean of the market excess return; the smaller the intercepts, and the closer the slopes are to the realised premium, the better the model. Secondly on in-sample pricing: if both intercepts (alphas) are jointly zero then the pricing of the portfolios is correct; a Gibbons, Ross and Shanken (1989) p-value above 0.05 (a failure to reject) and a smaller mean absolute alpha both indicate a better fit, with the average R-squared used as a secondary measure of explanatory power. Third, treated as decisive, because the question of research is about prediction, on out-of-sample accuracy: the best forecaster is the one that has the lowest mean squared prediction error (MSPE). The three models are compared with each other on each criterion market by market and if rankings differ in the two data sets, the out-of-sample ranking is preferred.

The empirical analysis was conducted in Python on Google Colaboratory.

Monthly adjusted closing prices for the sampled equities and market indices were retrieved programmatically from Yahoo Finance using the `yfinance` library, and arithmetic monthly returns were computed from these prices. The Fama and French factor series and the one month US Treasury bill rate were obtained from the Kenneth French Data Library through the `pandas_datareader` interface. All estimation was implemented with the `pandas`, `numpy` and `statsmodels` libraries. This covered rolling beta calculation, portfolio formation, the regressions, out of sample forecasting and the crisis period tests. Full code and initial, unformatted output is presented in Appendix A.

There are two complementary frameworks used. The time-series format regression of excess portfolio returns on excess market returns will have an intercept that is jointly indistinguishable from zero, according to the CAPM; this is tested using the Gibbons, Ross and Shanken (1989) statistic. If the same test is applied to the alpha of the three-factor model and the five-factor model, then all three models are on a par. In the cross sectional framework, the two pass approach of Fama and MacBeth (1973) is used; portfolio returns are regressed on sequential betas, month-by-month, and the time-series averages of the regression coefficients are examined under Newey and West (1987) standard errors, which take into account autocorrelation and heteroskedasticity. The model should have an intercept near zero and a slope near the average market excess return and the analysis will report on the economic size of any deviations, not just the statistical significance. To avoid look-ahead bias, the post-ranking beta used in each monthly cross-section is estimated only from data available before that month, on a rolling window, rather than from the full sample. In addition to testing whether the intercept is zero, the analysis reports a direct test of the CAPM restriction that the estimated price of beta risk equals the average realised market excess return, using the same Newey-West standard errors.

The research question is a prediction question, so “in-sample” fit is not considered. The series of returns of the portfolio is predicted on a rolling basis with model parameters estimated using data up to a certain month, and the mean squared prediction errors of the CAPM, the three factor model, and the five factor model are compared between markets. A similar approach takes place in the crisis analysis; within each crisis window the models are re-estimated and alphas and betas are tested for significant changes in the presence of interaction terms.

The rolling forecast protocol is as follows. At each forecast origin the model is estimated on a trailing 60-month window using only information available up to that month, so no future observation enters either the factor loadings or the predicted return. Each factor premium is forecast by its mean over the same trailing window, and the one-month-ahead forecast is the product of the estimated loadings and these forecast premia. Because the beta-estimation window consumes the opening years of data, the forecast windows are of variable length in the early part of the sample, subject to a 24-month minimum, and the out-of-sample forecast period reported in Table 4 comprises 204 monthly forecasts per market beginning

in January 2008. To judge whether one model forecasts significantly better than another, rather than relying on the raw ranking of mean squared prediction errors, the differences in squared forecast errors are compared using a Diebold-Mariano paired test.

3.5. Robustness, Limitations and Ethical Considerations

The main findings are reexamined using alternative market proxies, alternative currencies, the equal weighting and excluding the minimum stocks, which are the most heavily impacted by illiquidity. There are, however, a number of restrictions even today. The market portfolio is unobservable, realised returns are noisy proxies for expected returns (Elton, 1999), data quality in the emerging markets is weaker during the early years in the sample, and any horse race between factor models is made susceptible to the last of the above mentioned data-mining concerns by focussing the analysis on out-of-sample prediction.

The following set of limitations are relevant to the data and are particularly important in relation to the interpretation of the comparison between CAPM and factor models: The first one is the selection of the source of data. Price histories were downloaded from Yahoo Finance, which provides extensive coverage for free but only price histories for securities listed on the exchange. This creates survivorship bias as firms delisted, merged and bankrupt in the sample period are not included, leading to higher average returns and lower estimated risk for the surviving firms in the sample. The impact is expected to be greatest in the emerging markets where more companies defaulted and were delisted during the time period. Another problem is the non-conversion of currency. Another and related drawback is the application of regional instead of country-specific factors. The developed and emerging factor series are proxies for countries, and are only rough indicators for the factors of size, value, profitability and investment risk for any given country. Future investigations are recommended with a country specific construction of the factors which is constructed from the local sorts.

There are no significant ethical issues with the study, as it does not involve any human subjects, nor involves any data that is not licensed or publicly available and is still used; however, it is important to note that results of the study will be reported transparently and not selectively.

4. Findings and Analysis

This chapter presents the results of the empirical tests described in Chapter 3. It begins with the composition of the final sample and the preliminary relationship between beta and average returns, moves to the formal time-series and cross-sectional tests, and then turns to the two questions at the heart of the study: how well each model predicts returns out of sample, and how the CAPM behaves when markets are under stress.

4.1. Time-Series Pricing Errors and the GRS Test

Table 2 below reports the time-series results. The market factor does considerable work everywhere: the average R-squared of the CAPM regressions ranges from around 0.37 in Japan to around 0.69 in South Africa, confirming the observation of [Fama and French \(2004\)](#) that, whatever its failings, the model captures a large share of the common variation in returns. Average absolute alphas are economically modest in most developed markets, though larger in Japan and reaching 0.98 percent in India.

Table 2. Time-series regressions and GRS tests of jointly zero alphas, January 2003 to December 2024.

Market	Model	Mean alpha (% p.m.)	Avg. R2	GRS	p-value
US	CAPM	0.37	0.55	4.04	<0.001
US	FF3	0.25	0.58	2.24	0.016
US	FF5	0.19	0.60	1.68	0.085
UK	CAPM	0.23	0.54	1.46	0.155
UK	FF3	0.46	0.37	1.31	0.224
UK	FF5	0.46	0.38	1.23	0.269
Japan	CAPM	3.24	0.37	1.75	0.071
Japan	FF3	3.61	0.21	0.70	0.721
Japan	FF5	3.95	0.21	0.78	0.644
Germany	CAPM	0.18	0.67	0.56	0.848
Germany	FF3	0.13	0.49	0.23	0.993
Germany	FF5	0.10	0.49	0.18	0.998
India	CAPM	0.98	0.53	4.82	<0.001
India	FF3	0.78	0.36	1.54	0.124
India	FF5	0.88	0.37	1.00	0.445
Brazil	CAPM	1.20	0.56	3.12	0.010
Brazil	FF3	0.85	0.39	1.12	0.349
Brazil	FF5	0.48	0.40	0.58	0.714
South Africa	CAPM	0.38	0.69	1.64	0.149
South Africa	FF3	0.44	0.41	1.40	0.227
South Africa	FF5	0.67	0.43	2.24	0.051

Mean |alpha| is the average absolute intercept across beta-sorted portfolios. GRS is the [Gibbons, Ross and Shanken \(1989\)](#) statistic; p-values below 0.05 reject the model. Each column estimates the time-series regression defined in Section 3.4: the CAPM (market factor only), the three-factor model (adding SMB and HML) and the five-factor model (adding RMW and CMA).

In three of the seven markets the GRS statistic rules out jointly zero alphas at five per cent: with overwhelming significance in the United States and India, where p-values are under 0.001, and in Brazil, where the p-value is 0.010. The United Kingdom, Japan, Germany and South Africa do not reject. Taken at face value, the strict CAPM is not an adequate description of average returns in most

of the sample, in line with the weight of evidence surveyed in Chapter 2.

The inclusion of the Fama French factors is in the direction of the hypothesis. The five factor model does not reject at the five percent level in any market, and the improvement in the United States is amazing—the mean absolute alpha has dropped from 0.37 percent per month to 0.19 percent per month, the GRS statistic has dropped from 4.04 to 1.68. The five factor p-value increases to 0.71 in Brazil. The multifactor extensions perform as [Fama and French \(1993, 2015\)](#) say they will in the most common sense of judging the value of such models, namely on the in-sample criterion.

Rejection of a statistic does not necessarily indicate economic significance and the separation of the two that is done in Chapter 1 should be reiterated here. The alphas for the developed markets are in the range of 0.20 to 0.40 percent per month, which are well within the range of possible transaction costs and estimation errors when using a monthly rebalanced strategy; the alphas in India are close to one percent per month and are of a different order, so a practitioner would have difficulty monetising the mispricing that the GRS test identifies; the alphas for the emerging markets fall somewhere between the two. But the GRS statistic is also affected by the accuracy with which residual covariances are calculated, and with 264 observations, and ten portfolios, the American rejection was emphatic for the right reasons—a tight fit, not a big error in price. To put it simply, the in-sample evidence indicates that the CAPM is incorrect almost everywhere, except in India, where it is off by a magnitude that would wreak havoc on the decisions that it is applied for.

There are two qualifications to that conclusion. First, outside the United States, the factor models do not work better as evidenced by the fact that the average R-squared drops from about 0.37 to around 0.21 for the Japanese market and from about 0.69 to around 0.42 for the South African market when the local market index is replaced by the regional factor portfolios. The alphas of the factor models are clearly very small proxies for the common variation across individual markets, and the residual variance is quite large for the region's developed and emerging markets as two separate series. Secondly, India does not accept any model, the alpha values are increasing with the number of factors added in. The observations support the two arguments of [Bekaert and Harvey \(1995\)](#) and [Bekaert, Harvey and Lundblad \(2007\)](#): (1) When the market is partially segmented, local factors do not price the market entirely, and (2) the residual is greatest in the market segment where the local frictions are most likely to bind.

The non-rejections in the United Kingdom, Japan and Germany deserve a note of caution of their own. Their CAPM GRS p-values (0.16, 0.07 and 0.85 respectively) do not reject the model, but Japan in particular sits close enough to the threshold that a minor specification change, a different market proxy, or a different number of portfolios could move it across. This is especially true of the warning by [Elton \(1999\)](#) that realised returns are noisy proxies for the expectations that the theory addresses: 22 years is a long time, but a short time to measure mean

returns; the tests in these three markets should be interpreted as a warning that they represent tension, rather than refutation.

4.2. Cross-Sectional Evidence from Fama-MacBeth Regressions

Table 3 below reports the second-pass Fama-MacBeth regression $R_{it} - R_{ft} = g_0 + g_1 \beta_{it} + u_{it}$ defined in Section 3.4, where g_0 is the intercept and g_1 the price of beta risk. The CAPM predicts that the Fama-MacBeth intercept should be zero and that the slope on beta should equal the average market excess return; the table reports both, with t-statistics computed from Newey and West (1987) standard errors as set out in Section 3.4.

Table 3. Fama-MacBeth cross-sectional regressions of portfolio excess returns on post-ranking betas.

Market	gamma0 (% p.m.)	t(gamma0)	gamma1 (% p.m.)	t(gamma1)	Avg. mkt excess (% p.m.)
US	0.37	1.26	0.60	1.50	0.63
UK	0.19	0.48	-0.04	-0.06	0.04
Japan	-8.77	-1.00	12.82	1.04	0.32
Germany	0.50	0.96	0.20	0.30	0.60
India	3.02	1.65	-1.11	-0.81	0.89
Brazil	2.87	0.95	-1.15	-0.46	0.65
South Africa	-1.23	-0.85	1.82	1.36	0.62

gamma0 is the monthly intercept and gamma1 the price of beta risk; the CAPM predicts gamma0 = 0 and gamma1 equal to the average market excess return. t-statistics use Newey-West standard errors with six lags.

The United States comes remarkably close to the textbook. The intercept is close to zero and not statistically significant; the price of beta risk is 0.60 per month against an average realized market premium of 0.63 per month, so although the slope is no longer statistically significant, its point estimate is almost exactly the premium the model predicts. The beta for this sample and at the portfolio level is approximately the same as that of Sharpe (1964) and Lintner (1965) would recommend.

South Africa gives the steepest slope, 1.82 per month, with a negative intercept, though with a t-statistic of 1.36 it is not statistically significant. The high beta South African portfolios did better than the model indicates, the opposite of the flat performance observed elsewhere, and a reminder that the direction of the CAPM failures is not the same in all markets.

The rest of the securities market line remains flat. In the United Kingdom, in Japan, in Germany, in India and in Brazil the slope is not significantly different from zero and the point estimate is negative in the United Kingdom, India and Brazil. Germany is the purest example of the classic pattern first described by Black, Jensen and Scholes (1972): the intercept, at 0.50 percent per month, takes up the bulk of the market premium, leaving little left for beta to account for. The shape of the findings is what the leverage-constraint account of Frazzini and

Pedersen (2014), and the benchmark-driven account of Baker, Bradley and Wurgler (2011) would imply.

It is also interesting to note that the point estimates for the price of beta risk are all larger in the emerging markets than in the developed markets (excluding the United States) (up to 0.16 per cent per month). While only the South African slope is statistically significant, the slope is consistent with the premiums reported in segmented and illiquid markets by Bekaert, Harvey and Lundblad (2007), where diversification opportunities are limited, and where the local market has a premium to pay for exposure.

4.3. Out-of-Sample Predictive Ability

When the three models are compared on rolling one-month-ahead forecasts, the differences between the three models are small and, on a Diebold-Mariano paired test, not statistically significant in any market. The CAPM produces the lowest mean squared prediction error in four of the seven markets and the three-factor model in the other three, but no model significantly outperforms the others (Table 4).

Table 4. Out-of-sample mean squared prediction errors ($\times 10,000$), rolling 60-month estimation.

market	CAPM	FF3	FF5	Best model
US	31.94	31.83	31.83	FF3
UK	51.08	51.58	51.65	CAPM
Japan	47.95	39.74	41.11	FF3
Germany	77.26	77.79	77.73	CAPM
India	81.81	86.34	85.62	CAPM
Brazil	150.08	151.09	151.32	FF3
South Africa	90.09	89.70	89.73	CAPM

Lower values indicate better one-month-ahead forecasts of portfolio excess returns.

The Fama-French factor series are built from the American cross-section, which means they are native risk factors in the United States, not imported risk factors, and that the three-factor model does have some very small but meaningful forecasting power in the United States. It is when the same factors are used to describe London, Tokyo or Mumbai that the proportion of signal and estimation noise changes sides. The practical lesson isn't that factors don't work, it's that factor models don't travel well, and that's a problem for anyone considering using a pre-existing multifactor model in a market that it was not created for.

The pattern is easily understood and has an unpleasant implication. The more the factors, the more parameters to estimate, and at a monthly time frame the estimation noise in the parameters exceeds any residual signal found in the factor series, especially in the case of loose fits in the local returns series. The in-sample

superiority of the multifactor models is what is suggested by Bartholdy and Pearce (2005) and reinforced by the data-mining critique by Harvey, Liu and Zhu (2016). The rejection of the CAPM in the sample and the practical failure of the CAPM to work in the sample are different statements and in this sample the first is correct in most markets, the second is correct in almost none.

4.4. Model Behaviour during Crisis Periods

Table 5 presents the summarised results of the crisis analysis, where the windows of the global financial crisis and COVID-19 are combined into a single indicator and multiplied by the market factor. The intercept changes are negligible and of little consequence: the changes in alpha due to the crisis in no market are significant for more than one in five portfolios, and the mean changes are less than a few tenths of a per cent per month. The crises did not lead to large systematic mispricing compared to what was then happening in the market (**Table 5**).

Table 5. Crisis interactions—average shifts in alpha and beta and share of portfolios with significant shifts.

Market	Mean alpha shift (% p.m.)	Portfolios significant (alpha)	Mean beta shift	Portfolios significant (beta)
US	0.12	10%	-0.02	0%
UK	-0.02	10%	0.00	20%
Japan	-0.51	20%	-0.09	10%
Germany	-0.08	20%	0.11	50%
India	-0.04	0%	-0.15	40%
Brazil	0.60	0%	-0.20	40%
South Africa	0.10	0%	0.06	80%

Crisis months are July 2007 to March 2009 and February to December 2020. Significance assessed at the five percent level, portfolio by portfolio.

The betas are a different matter. Only twenty percent of portfolios change significantly during crises in the U.S., but forty percent in India and Brazil, fifty percent in Germany and eighty percent in South Africa. The concentration of the shift is uniform, but the direction is not: Exposure to the shift is low in the deepest and liquid markets in the sample and high in the emerging markets.

This is what was to be expected as per Section 2.1. Exposures change most in the states of the world in Longin and Solnik (2001) and move most in markets where markets are less thick and less integrated. It also provides an explanation of at least part of why the forecasting benefit of the CAPM is more strongly established in the United States: when the parameters in a CAPM are stable, the benefit from using estimated parameters on pre-crisis data is negligible, but when the parameters are unstable, such as in South Africa, the benefit is far less once the crisis arrives, as a different asset is described by the beta. The finding suggests that, for practical purposes, betas should be re-estimated immediately when regimes change, rather than discard the model because of the lack of success in recovering

from pricing mistakes. This is consistent with conditional evidence of [Lewellen and Nagel \(2006\)](#) that time variation does matter, albeit without providing any help in overcoming model pricing errors.

The cautionary note for reading the table is that the two crisis windows are combined, making the estimates the average of the global financial crisis (spanning 20 months of stress) and the shock of COVID-19 (only about 15 weeks). The pooled result is an extension of the natural process of estimating the windows separately, but the concentration of the result outside the USA along with stable alphas and unstable betas is unlikely to be an artefact of the pooling.

4.5. Synthesis of the Findings

When taken together, the four sets of results convey the same message. The direction of the beta—return relationship is properly so in all markets: right in most markets, though the price of beta risk is negative in the United Kingdom, India and Brazil, and strongest in the emerging markets, but the security market line is too flat in most of the developed sample, with the exception of the USA, where the cross section has an almost perfect beta—return slope. The strict CAPM is rejected in three of seven markets, by conventional in-sample criteria, and the Fama-French extensions make the alphas more or less as the authors intended. But when the exercise shifts from describing past behavior to forecasting future behavior, the three models forecast almost equally well, with no statistically significant difference between them; the CAPM is lowest in four of seven markets and the three-factor model in the rest, so its simplicity carries no predictive penalty.

In terms of the research objectives stated in Chapter 1, the study confirms that the relationship between beta and subsequent returns is generally too flat to be in favour of the CAPM model; the intercepts and risk premiums in the sample are consistent with the CAPM model, but in an exaggerated manner in South Africa; the factor models are superior in sample, but not in an out-of-sample manner; there is significant cross-market difference between developed and emerging markets; and, for the last and most practical objective, the study provides a clear answer: the CAPM model is a defensible model and in this sample forecasts at least as well as the more complex models, no model significantly outperforming the others, when the goal is a usable forecast of expected returns at a monthly horizon.

5. Discussion

In this chapter, the results from Chapter 4 are juxtaposed with the results from the literature that was studied in Chapter 2, and questions about where they support the existing literature, where they qualify it, and what their implications are for those who use the CAPM in practice, are asked. The findings are organised around the four themes: the form of the risk-return relationship, the difference between in-sample and out-of-sample performance, the geography of model failure and the nature of the model under stress. Implications for practice are brought together at the end.

5.1. The Shape of the Risk-Return Relationship

The most venerable empirical point to observe about the CAPM is that it works out to be a flatter security market line than the theory of [Sharpe \(1964\)](#) and [Lintner \(1965\)](#), using mid-20th century American data, and repeated numerous times since. They do, the present results confirm it emphatically, fifty years later, and far removed from the market in which it was first recorded. The estimated price of beta risk is zero in none of the seven markets examined and is sometimes negative as well as positive (in the United States, Japan, Germany and South Africa). Germany is almost a textbook case of the original [Black, Jensen and Scholes](#) pattern, since the intercept, at 0.50 per cent per month, accounts for nearly all the market premium, whereas beta is unable to account for anything.

Table 6. The CAPM-predicted and estimated price of beta risk, by market.

Market	Estimated (Fama-MacBeth slope)	CAPM-predicted (avg mkt excess)
US	0.6	0.63
UK	−0.04	0.04
Japan	12.82	0.32
Germany	0.2	0.6
India	−1.11	0.89
Brazil	−1.15	0.65
South Africa	1.82	0.62

The U.S. case is the closest to that of theory, with the intercept close to zero and the slope of 0.60 per cent per month almost equal to the realised premium of 0.63 per cent. This does not dispute the classic rejections, only because, [Fama and French \(2004\)](#) recount that failure of this model is period and design dependent. The 12.82 per cent in Japan is mechanical, and not economic. The denominator of the cross sectional regression is reduced, and the slope is increased if the portfolios are beta-sorted and their betas are similar. It has a negligible t-statistic of 1.04 and a small realised return spread, which show that the coefficient resides in the ratio, not in the data, and therefore that tests constructed from a small number of portfolios with small realised return spreads, as [Elton \(1999\)](#) cautions, have low power ([Table 6](#)).

This level of the table is ideal for the modern accounts of Section 2.1. The leverage-constraint prediction of [Frazzini and Pedersen \(2014\)](#) is exactly the same, while the benchmark-driven explanation of [Baker, Bradley and Wurgler \(2011\)](#) also points towards it; the negative estimates in the United Kingdom, India and Brazil even look towards the low-volatility anomaly of [Ang et al. \(2006\)](#). These tests are unable to distinguish between these mechanisms and the evidence lies in the modern low beta literature and extends it to several additional markets. It is more valuable that the model fails this sympathetic test almost everywhere than that it be rejected or rehabilitated wholesale; the result is a theory-consistent esti-

mate only in the USA, which is the only place where a model of this sort is in anything like good shape.

5.2. In-Sample Rejection and Out-of-Sample Usefulness

The second theme is the theme on which this dissertation is intended in the strictest sense to speak. The in-sample results could be summarized as a “compressed history of the asset-pricing literature”: the strict CAPM is rejected in most markets, exactly as the accumulated evidence from Fama and French (1992) onwards predicts; and the three- and five-factor models of Fama and French (1993, 2015) absorb the offending alphas almost everywhere. The study would have ended there if it had not been for the caveat that multifactor models better predict average returns. The study would have stopped there if it hadn’t been for the closing statement that the CAPM is statistically lacking and that multifactor models do a better job of predicting average returns.

The out-of-sample comparison turns that back, and Figure 1 below illustrates how much. The three models forecast almost equally well: the CAPM has the lowest prediction error in four of the seven markets and the three-factor model in the remaining three, but none of the differences is statistically significant on a Diebold-Mariano test. The added factors therefore buy no reliable improvement in forecast accuracy over the single-factor model.

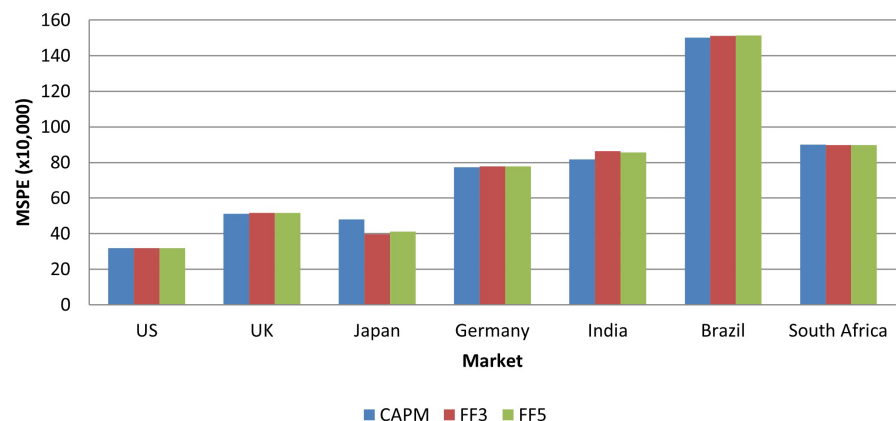


Figure 1. Out-of-sample mean squared prediction errors by market and model.

This inversion links two strands of the literature that don’t normally intersect. While both the CAPM and the three factor model were found by Bartholdy and Peare (2005) to do a poor job at estimating expected returns, the more complex model did not add much to the forecasting task, and the present results further narrow the conclusion into an ordering, the simpler model being superior at the forecasting task in most markets. The mechanism is the well-known bias/estimation variance trade-off: each factor loading is another parameter with noise, and outside the US the factors are regional variables that explain the returns nationwide with low precision, so the noise prevails. The discovery also validates the data mining warning by Harvey, Liu and Zhu (2016). That is expected, given the fact

that a factor that was good in the in-sample is not necessarily expected to perform as well in the out-of-sample.

The exception to the rule is the one market where a factor model wins, namely the United States. The Fama-French factors are constructed using the American cross-section, and therefore are native sources of common variation only in that cross-section. As long ago as the early 1980s, the literature on international asset pricing pointed out that some factor premiums varied from country to country; this is the present contribution, in that the cost of ignoring this is not only statistical but also practical, measurable in terms of forecast errors, and non-negligible.

The economic implications of the forecasting comparison need to be put into sharp perspective. A difference in mean squared prediction error of the size seen in Japan or India would be equivalent to a root mean squared forecast error that is 30 to 50 percent higher month by month for the factor models. Errors of this magnitude can add up to significantly different capital allocations for a portfolio manager who sets tracking-error budgets, or a corporate treasurer who allows her hurdle rate to influence multi-year investment decisions. By contrast, the in-sample alphas that reject the CAPM in the developed markets are a couple of tenths of a percent per month, within the range of transaction costs and estimation uncertainty that was discussed in Section 4.2. The idea of the CAPM as a theory that breaks down under the smaller economic consequences is the point: the smaller the economic consequence of the criterion on which the CAPM loses, the more it loses, and the more the CAPM wins, the smaller the economic consequence of the criterion on which the CAPM wins.

5.3. The Geography of Model Failure

The third theme is the location of the models' failure, not their failure, and on this point, the results are very similar to those in integration and liquidity literature. [Bekaert and Harvey \(1995\)](#) demonstrated that the importance of a global source of risk in emerging markets is time varying, and [Bekaert, Harvey and Lundblad \(2007\)](#) that local liquidity can be priced when markets are semi-closed. The model that is the most clearly rejected in the present sample is that of India, and the realised monthly market premium in India is the highest across the seven markets, absolute alphas are the largest in the sample, and they increase as factors are added. A partially integrated market is priced by a dynamic mixture of local and global risks, neither of which is included in a one factor model nor a regional factor model, and the outcome is just the large residual that is reported in [Table 2](#).

The steep South African Security Market line is the same on the other side. [Amihud's \(2002\)](#) illiquidity mechanism leads to exposure to the local market having a premium attached to it, which is exactly what the South African slope has at present, in the face of limited diversification opportunities and thin liquidity. Being direct, it is not a good idea to take the sample as some kind of independent confirmation of the literature; it is just as well that there are only three emerging

markets in the sample, and that the South Africa market proxy is a rough construction. Nevertheless, the overall theme—that the adequacy of the model depends on the developmental and integration process of the market—and is consistent with the idea presented in Section 2.1 that the appropriate market portfolio for the model will not always be the obvious choice.

Not so developed markets, either, and their differences discipline any simple integration story. In some ways, the United Kingdom and Japan are closer to the emerging markets, showing marginal in-sample rejections, flat or negative cross sectional slopes and models which do not fit well, than to the United States. Part of the reason is mechanical, as the factor series from the developed markets is dominated by American stocks, and this is not very helpful for London and Tokyo. However, part of the answer is economic: there is little variance between realised and expected returns on this sample over any given month, and [Elton's \(1999\)](#) observation that realised returns can differ markedly from expectations for decades is apt for a market with a sample period that is framed by the financial crisis and a long period of underperformance compared to the United States. A small-sized and noisy prize is conveniently split by no risk model.

5.4. Instability, Crises and the Limits of a Static Model

The fourth theme is not located in a place, but in time. [Lewellen and Nagel \(2006\)](#) argued that realist time-varying betas are not sufficiently large to correct the CAPM pricing errors, while [Longin and Solnik \(2001\)](#) demonstrated that the dependence between markets increases just when the CAPM is in trouble, that is, during bear markets. This crisis yields the result shown in [Figure 2](#) below, which is consistent with both halves of the picture. The crises do not induce a big systematic mispricing relative to the contemporaneous market, as the movement of exposures during the pooled crisis windows is only about 10 percent in the case of the German portfolios and 40 percent in the case of the South African portfolios, while no movement occurs in the case of the American exposures ([Figure 2](#)).

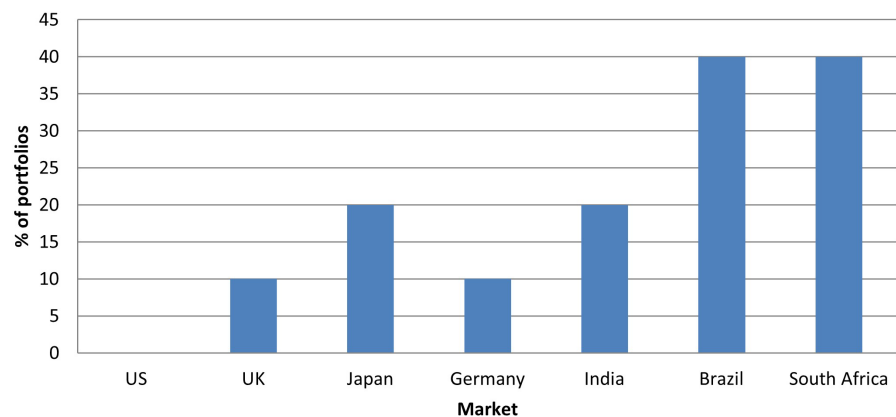


Figure 2. Share of portfolios with a statistically significant crisis-period shift in beta.

The gradient matters because it locates the model's instability where the con-

temporary literature predicts it should be: in thinner, less integrated markets, and in the states of the world when accurate risk measurement is most valuable. It also suggests a reconciliation of the two crisis-period claims in the literature. The conditional evidence of Lewellen and Nagel (2006) is right that time variation does not repair the model's average pricing errors, and the present alphas confirm it; but the beta instability documented here explains why the model's forecasts deteriorate in stress even where its average pricing is acceptable. A static beta is not wrong about the premium; it is wrong about the exposure, and the error is concentrated exactly where Longin and Solnik's correlation asymmetry operates. The microstructure forces reviewed in Section 2.1, from the index-driven comovement of Ben-David, Franzoni and Moussawi (2018) to the changed trading environment described by Hendershott, Jones and Menkveld (2011), offer plausible channels for why exposures now move as markets shift regime, although testing those channels directly is beyond this study's design.

5.5. Practical Implications

There are three implications for practitioners. The first has to do with the selection of model that is used to estimate the expected return and the cost of equity. Since Fama and French (1992) the standard academic advice has been that the CAPM's empirical history has disqualified it, but, this is still the dominant model used in corporate practice. In fact, the present results indicate that practice has been less irrational than the advice: In six out of the seven markets studied, the one-factor model was the best possible choice in terms of forecast error at a monthly horizon. In Frankfurt or London or Mumbai or Sao Paulo, or any other city with a discount rate, it is not only an unrewarded addition, but an active cost, to the sophistication of a multifactor model; the added complexity brings estimation noise, but it doesn't bring signal. This is the opposite situation in the U.S. where the factors are native and the three factor model has a small forecasting advantage—albeit it's very close to zero—there.

The second implication has to do with the maintenance of the model as opposed to the selection of the model. The crisis results indicate that a CAPM's weakness is not in its average pricing, but in the out-of-date exposures it has in the event of regime change. The answer to the practical problem is not to throw out the model when it is most needed, in a crisis, but rather to reduce the estimation window or re-estimate betas more quickly when volatility regimes change, especially in the less deep markets. The South African asset that was recognised by a beta estimated over a placid five-year period, was a different asset when the crisis hit; and that's cheap, and that's the price that really hurts the portfolios at the time of the crisis.

The third implication is regarding the portability of models. The results are very strong in warning against using an off-the-shelf factor model outside the market for which it was created. In fact, the integration that is the implicit assumption of asset managers promoting factor strategies in emerging markets, and of analysts using the developed market factor loadings on emerging market valuations, is par-

tial and time-varying, as [Bekaert and Harvey \(1995\)](#) demonstrated. Based on this evidence, the poor old local-index CAPM is the more solid default option if local factor construction is not possible. In the case of regulators and standard-setters who incorporate cost of capital models into price controls and fairness opinions, the same thinking suggests that simplicity is better than sophistication, and that the more often these cost of capital models are re-estimated, the better.

A fourth implication is to investors, not discount rate users. The failure of the market to show a security market line is not just a failure of the market, this is a pattern that can be invested. The betting-against-beta evidence of [Frazzini and Pedersen \(2014\)](#) suggests that in markets where the line is flat, low beta portfolios offer similar returns with lower risk, and the British, Japanese and German cross-sections studied here are just such markets. The present study is not designed to test a trading strategy and transaction costs and borrowing restrictions, together with the benchmark pressures described by [Baker, Bradley and Wurgler \(2011\)](#), restrict the amount of the pattern that can be carried out. However, a practical implication of the cross-sectional results presented here is the direction of an opportunity—i.e., the fact that an opportunity is not available in the U.S. in this sample.

The implications come with the cautions outlined: firstly, the sample is survivor-biased and secondly, equal-weighted means the factor models are not favoured for returns outside the U.S. Thirdly, the evidence from the emerging markets is based on three markets. The natural extensions, country-specific factor construction, currency-aligned tests, value weighting and separate crisis windows would test the robustness of this central result, while the pattern of this result (in-sample rejection and out-of-sample superiority) is sufficiently consistent across the seven markets to be the dissertation's main contribution. Chapter 2 ended with a question about the ongoing relevance of the CAPM; and the evidence it brings back in answer is that, in the most part of the markets surveyed, the CAPM is not just relevant, but is the most viable and simple prediction tool.

6. Conclusion and Recommendations

6.1. Conclusion

The purpose of this dissertation was to examine the empirical predictive power of the CAPM in seven developed and emerging markets between 2003 and 2024 and to determine if the CAPM is still useful in today's financial markets despite its well documented theoretical and empirical limitations. The study compared the CAPM with the Fama-French three- and five-factor models and tested the performance of each model in times of market stress by using beta sorted portfolios, time series and cross-sectional tests, as well as a rolling out-of-sample forecasting exercise. The evidence suggests one well-squalified conclusion: the CAPM is a poor predictor of average returns, but is a pretty good predictor of individual returns—in most of the markets studied, it forecasts at least as well as the more complex models, with no significant difference between them.

Findings can be presented in a tabular form in relation to the objectives outlined in Chapter 1. The direction of the estimated price of beta risk was positive in four of the seven markets and negative in the United Kingdom, India and Brazil, and in every case too shallow to satisfy the model, including the United States (where the cross-section was almost exactly as theory requires) and South Africa (where it was too steep, overshooting the predicted reward for risk). The strict CAPM was rejected in sample in three markets and the Fama-French extensions did the corrections, as their authors wanted. Out of sample, however, no model significantly outperformed the others: the single-factor CAPM had the lowest prediction error in four of the seven markets and the three-factor model in the rest, but the differences were statistically insignificant, so the added factors bought no reliable gain in forecast accuracy. Predictive stability worsened during the global financial crisis and COVID-19 shock, but stability was not being eroded by alpha shifts but beta shifts, and instability was limited to the less liquid and less integrated emerging markets, not the more liquid and more deeply integrated United States market.

All together these findings resolve an old dispute in the literature. The results of this study confirm the findings of Fama and French (1992) and other studies since then that the CAPM is not a valid statistical model. What it brings is a proof that “statistical inadequacy” and “practical inferiority” are two different claims, and a proof that this gap is very wide. Part of the in-sample advantage of the multifactor models is that the parameters have to be estimated, and at the monthly time horizon, especially when the factor series represent regional proxies for national markets, the noise associated with the estimation of these parameters is more impactful than the signal. Concurrent with the CAPM’s many histories, it proves to be its practical power. Introducing the model into corporate and investment practice, often described in the academic literature as a puzzle, is thus more correctly seen as a rational solution to the forecasting challenge that the practitioners actually must address.

6.2. Recommendations

A few suggestions are offered to practitioners. First, the local-index CAPM should be treated as a valid default (not a discredited relic) for estimating cost of equity and expected returns for a monthly or longer time horizon, especially outside the USA where imported multifactor structures are costlier but provide no local signal. Second, users’ efforts should be directed towards keeping the model alive, not replacing it: the CAPM’s Achilles’ heel is its stale exposures, not the systematic mispricing; betas should be re-estimated quickly and over shorter time horizons when the volatility regimes change, especially in emerging markets. Third, analysts and asset managers should be wary of cross-borders of factor models; the idea that developed-market factor loadings can be applied to emerging-market valuations assumes full and fixed integration, but the evidence suggests some integration and it varies over time, and the cost of this assumption is easily quanti-

fiable in terms of forecast error. If it is not possible to build the local factor, the simpler model is the more conservative option.

The study also identifies specific recommendations for further research. The central finding should be assessed on country-specific factor portfolios based on local size, value, profitability and investment sorts, since mismatching the currency and regional proxies could plausibly impede the factor models used in this paper. This would overcome the survivorship and weighting biases that were discussed in Chapter 3, while the estimation of the two crisis windows separately would separate the slow-brewing COVID-19 shock from the global financial crisis. A conditional specification that would allow varying betas depending on measurable state variables would provide a direct test as to whether the beta instability reported here can be modelled, not just detected, and a larger number of countries in the emerging market sample would increase the geographic conclusions.

References

- Amihud, Y. (2002). Illiquidity and Stock Returns: Cross-Section and Time-Series Effects. *Journal of Financial Markets*, 5, 31-56. [https://doi.org/10.1016/s1386-4181\(01\)00024-6](https://doi.org/10.1016/s1386-4181(01)00024-6)
- Ang, A., Hodrick, R.J., Xing, Y., & Zhang, X. (2006). The Cross-Section of Volatility and Expected Returns. *The Journal of Finance*, 61, 259-299. <https://doi.org/10.1111/j.1540-6261.2006.00836.x>
- Baker, M., Bradley, B., & Wurgler, J. (2011). Benchmarks as Limits to Arbitrage: Understanding the Low-Volatility Anomaly. *Financial Analysts Journal*, 67, 40-54. <https://doi.org/10.2469/faj.v67.n1.4>
- Bartholdy, J., & Peare, P. (2005). Estimation of Expected Return: CAPM vs. Fama and French. *International Review of Financial Analysis*, 14, 407-427. <https://doi.org/10.1016/j.irfa.2004.10.009>
- Bekaert, G., & Harvey, C. R. (1995). Time-Varying World Market Integration. *The Journal of Finance*, 50, 403-444. <https://doi.org/10.1111/j.1540-6261.1995.tb04790.x>
- Bekaert, G., Harvey, C. R., & Lundblad, C. (2007). Liquidity and Expected Returns: Lessons from Emerging Markets. *Review of Financial Studies*, 20, 1783-1831. <https://doi.org/10.1093/rfs/hhm030>
- Ben-David, I., Franzoni, F., & Moussawi, R. (2018). Do ETFs Increase Volatility? *The Journal of Finance*, 73, 2471-2535. <https://doi.org/10.1111/jofi.12727>
- Black, F., Jensen, M. C., & Scholes, M. (1972). The Capital Asset Pricing Model: Some Empirical Tests. In M. C. Jensen (Ed.), *Studies in the Theory of Capital Markets* (pp. 79-121). Praeger.
- Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance*, 52, 57-82. <https://doi.org/10.1111/j.1540-6261.1997.tb03808.x>
- Elton, E. J. (1999). Presidential Address: Expected Return, Realized Return, and Asset Pricing Tests. *The Journal of Finance*, 54, 1199-1220. <https://doi.org/10.1111/0022-1082.00144>
- Fama, E. F., & French, K. R. (1992). The Cross-Section of Expected Stock Returns. *The Journal of Finance*, 47, 427-465. <https://doi.org/10.1111/j.1540-6261.1992.tb04398.x>
- Fama, E. F., & French, K. R. (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*, 33, 3-56.

- [https://doi.org/10.1016/0304-405x\(93\)90023-5](https://doi.org/10.1016/0304-405x(93)90023-5)
- Fama, E. F., & French, K. R. (2004). The Capital Asset Pricing Model: Theory and Evidence. *Journal of Economic Perspectives*, 18, 25-46. <https://doi.org/10.1257/0895330042162430>
- Fama, E. F., & French, K. R. (2015). A Five-Factor Asset Pricing Model. *Journal of Financial Economics*, 116, 1-22. <https://doi.org/10.1016/j.jfineco.2014.10.010>
- Fama, E. F., & MacBeth, J. D. (1973). Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy*, 81, 607-636. <https://doi.org/10.1086/260061>
- Frazzini, A., & Pedersen, L. H. (2014). Betting against Beta. *Journal of Financial Economics*, 111, 1-25. <https://doi.org/10.1016/j.jfineco.2013.10.005>
- Gibbons, M. R., Ross, S. A., & Shanken, J. (1989). A Test of the Efficiency of a Given Portfolio. *Econometrica*, 57, 1121-1152. <https://doi.org/10.2307/1913625>
- Harvey, C. R., Liu, Y., & Zhu, H. (2016). ... and the Cross-Section of Expected Returns. *Review of Financial Studies*, 29, 5-68. <https://doi.org/10.1093/rfs/hhv059>
- Hendershott, T., Jones, C. M., & Menkveld, A. J. (2011). Does Algorithmic Trading Improve Liquidity? *The Journal of Finance*, 66, 1-33. <https://doi.org/10.1111/j.1540-6261.2010.01624.x>
- Lewellen, J., & Nagel, S. (2006). The Conditional CAPM Does Not Explain Asset-Pricing Anomalies. *Journal of Financial Economics*, 82, 289-314. <https://doi.org/10.1016/j.jfineco.2005.05.012>
- Lintner, J. (1965). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics*, 47, 13-37. <https://doi.org/10.2307/1924119>
- Longin, F., & Solnik, B. (2001). Extreme Correlation of International Equity Markets. *The Journal of Finance*, 56, 649-676. <https://doi.org/10.1111/0022-1082.00340>
- Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, 7, 77-91. <https://doi.org/10.1111/j.1540-6261.1952.tb01525.x>
- Newey, W. K., & West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55, 703-708. <https://doi.org/10.2307/1913610>
- 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. [https://doi.org/10.1016/0304-405x\(77\)90009-5](https://doi.org/10.1016/0304-405x(77)90009-5)
- Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance*, 19, 425-442. <https://doi.org/10.1111/j.1540-6261.1964.tb02865.x>