Chapter 4: Empirical tests

Chapter summary: This chapter investigates the empirical questions of stochastic dominance (SD) efficiency and two-fund separation under SD conditions, and elaborates on their relation. To do this, we use the tests for stochastic dominance efficiency as developed in Chapter 2 to test for efficiency of the market portfolio for the US markets. It is demonstrated that the mean-variance inefficiency of the CRSP all-share index relative to several types of sorted portfolios can be explained by SD models. We also test for two-fund separations using the approach outlined in Chapter 3. Especially in case of the beta portfolios, SD efficiency tests provide a better explanation than the MV framework (based on tail risk not captured by variance). The TFS-test underlines the robustness of the specification for the size and reversal portfolios.

The concepts of SD efficiency (Chapter 2) and Two-fund separation (Chapter 3) are related. This not only follows from their respective definitions (see Chapter 1), but also from the impact of these questions on the representative investor assumption. Efficiency is a necessary condition for two-fund separation and two-fund separation is a necessary condition for aggregation; note however, that neither condition is also sufficient. This chapter focuses on an empirical examination of these conditions.

4.1 Efficiency, Two-fund separation, and Aggregation

In section 1.4 it was shown that there are potentially three ways to arrive at aggregation (defined as a situation wherein the investors choose their portfolios in such a way that their choices could be modeled as if a single 'representative' investor makes them). These were:

i. Complete markets. All risk is tradable and will be traded if preferences give rise to it. As an empirical explanation this way of arriving at aggregation offers little hope; normally the hypothesis of complete markets will be rejected out of hand.

ii. Homogeneity of preferences. If preferences are similar enough between individual investors, they will choose the same portfolios and market portfolio efficiency is obtained. We have seen in Section 1.4.3 that only a rather limited subset of utility specifications will lead to aggregation in this way (the requirements are actually stricter than mere aggregation) and hence support two-fund separation.

iii. Convexity of the efficient set. If utility and return distributions combine in such a way that the efficient set is convex, market efficiency would obtain; this without further assumptions on the set of tradable assets (such as completeness). We have not been able to derive exact conditions under which this line of argument could lead to aggregation (it may be impossible, as we only have negative proof).

The null hypothesis of efficiency and the null of TFS can both be related to this list:

First, if SD efficiency is rejected, all representative investor models can be rejected. In this situation, we would need to relax the assumptions behind single-period, consumption oriented, representative investor models in order to describe capital market equilibrium. Neither two-fund separation nor aggregation would be possible.

Second, if TFS is not rejected, no representative investor model *that conforms to the SD-criterion used* can be rejected. In other words, we would have to pick one of the above 3 ways of obtaining aggregation; or assume that while none of these 3 situations hold, the investors behaves almost as if one of them holds and the differences are such that the violation is not detectable. The latter option may seem simply a methodological oddity, but given the (infinite) range of utility functions that are possible it may very well be possible that investors do not conform to the rather tight restrictions imposed by the homogeneity conditions, yet produce a result that is hard to distinguish from it. Of course, mere chance is also included in this category - to be exact, this would be a Type II error (failure to detect violations of TFS). However, the TFS test is set up in such a way as to minimize this possibility (see Chapter 3).

However, depending on the SD order used, we could potentially rule out option (ii), as we do not have an SD order that imposes the same conditions as those which are necessary-and-sufficient for (ii) to hold. In other words; each SD criterion we have is broader, and hence should point to option (i) or (iii). Since completeness of the markets is dubious at best, not rejecting TFS would point to convexity of the efficient set, or a situation where none of the 3 options hold but the empirical situation is quite close to it. It would be a significant indicator of the usefulness of further research into the convexity of the efficient set under SD conditions.

Also, there are relevant relations between TFS and efficiency of different orders. Figure 1 illustrates this, this is an extension of the results given in Chapter 3. Combined, the concepts of SSD and TSD efficiency and their TFS counterparts yield five possible scenarios (summarized in the diagram below).

Firstly, if SSD-TFS applies, then the market portfolio is optimal for all well-behaved risk averse utility functions and hence all representative investor models apply. This form of robustness for the model specification is comforting given that there exist few prior arguments for selecting a specific model (for instance the mean-variance model or the mean-semivariance model). Of course, if SSD-TFS applies, then the market portfolio must also be SSD efficient. Furthermore, if SSD-TFS applies, TSD-TFS logically follows, since well-behaved utility functions under TSD are by definition elements of U_2. The same logic holds for any further subsets of U_3, such as 4SD and DARA-SD.

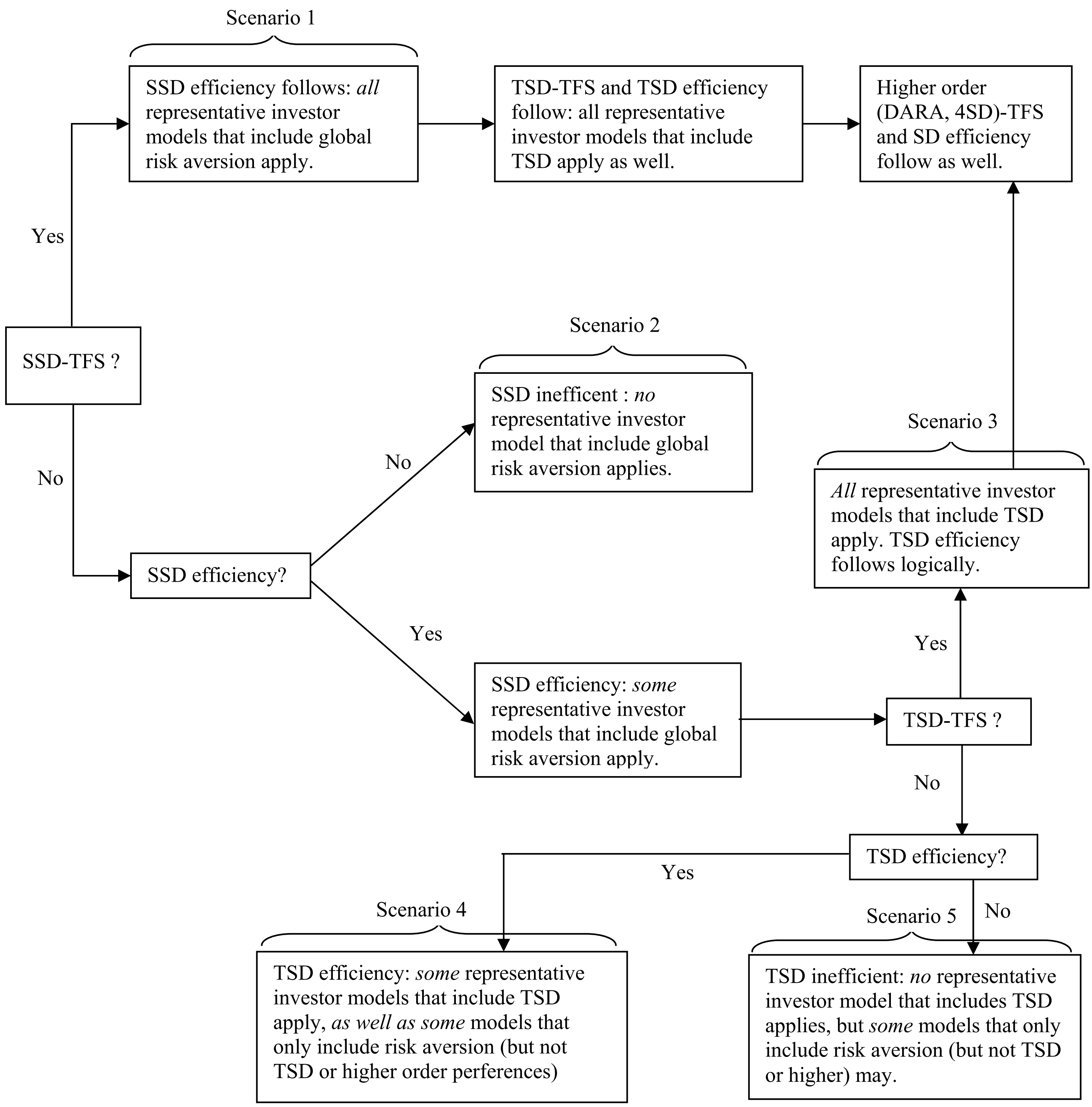
Secondly, if SSD-TFS does not apply, we might also find that the market is SSD inefficient. In this scenario, there is no utility function for which the market is optimal and hence no representative investor model incorporating (global) risk aversion applies. In this case, we need a more a more complex model to describe capital market equilibrium, for instance an intertemporal consumption-oriented model or a heterogeneous investor model. The reverse relationship is a sufficient one: SSD inefficiency always leads to rejection of SSD-TFS.

The third, fourth and fifth scenarios are all based on the case where SSD-TFS doesn't apply, yet the market is SSD efficient, meaning that only some representative investor models which assume risk aversion are valid. In the third case, TSD-TFS applies, meaning that all models with DARA are valid and that TSD efficiency logically follows. TFS and efficiency of higher orders also obtain.

The fourth and fifth scenarios arise when TSD-TFS doesn't apply, leaving only SSD efficiency and SSD and TSD efficiency respectively, but no TFS. In the former case, a subset of TSD models and the representative investor models that do assume risk aversion but not TSD (or higher orders) cannot be rejected. In the latter situation we are left only with models that do assume risk aversion, but are incompatible with DARA or higher order SD-rules. These last two situations are uncomforting given the uncertainty surrounding the correct specification and the possibility of specification error.

Figure 1.

Relations between SSD and TSD efficiency and Two-Fund Separation (TFS). The figure starts at SSD-TFS, and details the logical relations. As described in Chapter 3, in some cases it is possible to get empirical results that do not allow this schedule to be followed due to different measures of errors; the efficiency tests deflate the errors, while the TFS-tests inflate them, so we obtain a conservative approach for both tests.



4.2 Data and methodology

This chapter will investigate the properties of several sets of portfolios. In each case, we will use 10 portfolios sorted on a certain characteristic, usually associated with a certain 'anomaly'. The purpose is twofold: to see if these anomalies might be explained by a more general risk-measure (in line with the SSD and TSD concepts), and to see if the data supports TFS and aggregation. The latter issue is relatively new, to our knowledge no paper follows a similar course. The former issue has been researched extensively in a MV framework, but also using SD criteria. Examples of the latter are Van Vliet (2004), on which part of this chapter is inspired, and Post-van Vliet (2005). It should be noted that there is a rich body of literature that questions these type of empirical endeavours; Lo & MacKinley (1990) and White (2000) suspect 'datasnooping' – i.e. picking up accidental patterns – plays a non-negliable role, while the choice of the market-portfolio is also a recurring (and ultimately, unsolvable) problem, see for example Roll (1977). Finally, the word 'anomaly' should always be taken with a pinch of salt, as ultimately an anomaly only qualifies as such if a model cannot explain a fact; and there are almost infinite possibilities for explanation if one wishes to use conditional (Campbell and Cochrane, 2000) and intertemporal models (following Merton, 1973), let alone abandon rationality (Lakonishok et al, 1994). Campbell (2000) describes how wide the field of Asset Pricing has become. This empirical chapter has a far more modest scope, and will be based on a single-period, representative investor model using rationality and our SD approach.

The methodology employed has been discussed in Chapter 2 and 3 of this thesis. We will use the SSD and TSD efficiency tests (without restrictions) from sections 2.2 and 2.3, and the TFS test as described in 3.2. We will also use the GRS (Gibbons Ross Shanken, 1989) test for comparison and to make a link with the Capital Asset pricing Model (CAPM) and the MV framework in general. Since the evaluated portfolio is (a proxy for) the market portfolio, the GRS test is a direct test for the central prediction of the CAPM that the market portfolio is mean-variance efficient. In this respect, the SSD and TSD efficiency tests can be seen as tests for generalizations of the CAPM; single-period, portfolio-oriented, representative-investor asset pricing models that use a more general definition of risk than the CAPM does.

To illustrate the SD results, which often point towards 'tail-risk' (risk aversion – or even increasing aversion – in the domain of the very negative returns) we will also use the 'tail-beta' of one of our set of portfolios.[1] Tailbetas are useful to quantify the contribution of stocks to the tail risk of the market portfolio. Following Bawa and Lindenberg (1977), Price, Price and Nantell (1982) and Harlow and Rao (1989), we may measure this contribution by means of the second-order co-lower partial moment:

$$CLPM_{2,i} \equiv \frac{\frac{1}{T}\sum_{t=1}^{T}\max(m-\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\tau},0)(x_{i,t})}{\frac{1}{T}\sum_{t=1}^{T}\max(m-\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\tau},0)(\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\tau})} \qquad (1)$$

with the target rate m again set equal to –10%, (compare chapter 2). Interestingly, this "tail beta" equals the regular beta if the return distribution is normally distributed (see Bawa and Lindenberg, 1977). However, for skewed distributions, the measures may diverge.

The data used in this chapter comes from the CRSP files. Most portfolios are created by Prof. K. French[2]; the beta-sorted portfolios were created using software provided by Pim van Vliet. Descriptive statistics can be found in table I; a short description of the way the dataset is constructed is given below. Excess returns are computed from the raw return observations by subtracting the return on the one-month US Treasury bill from Ibbotson.

Our Dataset contains the following portfolios:

Beta sorted portfolios:[3] The sample period runs from January 1933 to December 2002 (monthly observations; T=840). The beta portfolios are constructed from the CRSP tapes. In December of each year, all stocks that fulfill our data requirements are placed in ten portfolios based on the previous 60-month betas. A minimum of 12 months of

---

[1] To conserve space, we only use this analysis for the beta-sorted portfolios, which show a clear distinction between the SD and MV results for the full sample.

[2] This data can be found on http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. The descriptions later on are adapted from this page as well.

[3] We thank Pim van Vliet for making the data available. This data can be found at his online datacenter: http://www.few.eur.nl/few/people/wvanvliet/datacenter.

return observations is needed for a stock to be included on formation date. Each portfolio includes an equal number of stocks.

Size sorted portfolios: The sample period runs from January 1927 to December 2002 (monthly observations; T=912). The portfolios are constructed at the end of each June using the June market equity and NYSE breakpoints. The portfolios for July of year t to June of t+1 include all NYSE, AMEX, and NASDAQ stocks for which there is market equity data for June of year t available. (See French, 2006)

Book-to-market ratio sorted portfolios: The sample period is January 1927 to December 2002 (monthly, T=912); portfolios are formed on Book Equity (BE) divided by Market Equity (ME) at the end of each June using NYSE breakpoints. The BE used in June of year t is the book equity for the last fiscal year end in t-1. ME is price times shares outstanding at the end of December of t-1. The portfolios include all NYSE, AMEX, and NASDAQ stocks for which there is a ME for December of t-1 and June of year t, and BE for year t-1. (See French, 2006)

Momentum sorted portfolios: This datasets contains observations from January 1927 to December 2002 (monthly, T=912). The portfolios are constructed monthly based on prior (2-12) return decile breakpoints, using NYSE, AMEX, and NASDAQ stocks with prior return data. To be included in a portfolio for month t (formed at the end of the month t-1), a stock must have a price for the end of month t-13 and a good return for t-2. In addition, any missing returns from t-12 to t-3 must be -99.0, CRSP's code for a missing price. Each included stock also must have ME for the end of t-1. (See French, 2006)

Long Term reversal sorted portfolios: This datasets contains monthly observations from January 1931 to December 2002 (T=864). The portfolios are constructed monthly based on prior (13-60) return decile breakpoints, using NYSE, AMEX, and NASDAQ stocks with prior return data. To be included in a portfolio for month t (formed at the end of the month t-1), a stock must have a price for the end of month t-61 and a good return for t-13. In addition, any missing returns from t-60 to t-14 must be -99.0, CRSP's code for a missing price. Each included stock also must have ME for the end of t-1. (See French, 2006)

Table I

Descriptive Statistics Benchmark Portfolios

The table shows descriptive statistics for the benchmark portfolios formed on market beta (Panel A) and the monthly excess returns of the CRSP index, as well as the EP, TP and LP test portfolios constructed for our simulations (Panel B). The reported kurtosis is the excess kurtosis.

Panel A : The CRSP market index and the test portfolios						
	Mean	Stdev.	Skewness	Kurtosis	Min	Max
Market	0.714	4.937	0.156	6.181	-23.673	38.172
EP	0.774	5.699	0.560	9.025	-28.020	47.953
TP	0.960	4.264	-0.022	4.571	-21.870	27.730
LP	0.960	5.721	1.361	9.156	-15.139	53.311
Panel B : The 10 benchmark portfolios based on beta						
	Mean	Stdev.	Skewness	Kurtosis	Min	Max
Low β	0.670	3.822	-0.754	5.230	-24.577	15.718
2	0.698	4.015	-0.018	3.926	-20.573	24.222
3	0.756	4.631	0.648	10.175	-25.003	41.292
4	0.659	4.832	0.255	6.269	-25.943	34.332
5	0.918	5.669	1.041	13.370	-29.333	55.762
6	0.833	6.094	0.592	8.279	-28.615	48.932
7	0.809	6.538	0.574	8.773	-32.573	53.842
8	0.768	7.470	0.774	9.264	-30.395	61.832
9	0.833	8.306	0.689	7.941	-36.583	64.262
High β	0.794	9.653	0.814	8.516	-37.133	83.692
Panel C : The 10 benchmark portfolios based on size						
	Mean	Stdev.	Skewness	Kurtosis	Min	Max
Small	1.501	10.282	3.443	31.896	-34.59	115.69
2	1.328	9.025	2.352	22.931	-32.94	94.87
3	1.284	8.065	1.785	18.224	-32.83	77.73
4	1.236	7.543	1.557	15.803	-31.47	66.67
5	1.212	7.287	1.236	13.898	-31.02	62.32
6	1.177	6.913	1.029	12.156	-31.15	54.94
7	1.158	6.564	0.851	11.405	-29.25	53.24
8	1.091	6.225	0.793	11.119	-30.80	52.53
9	1.044	5.952	0.731	11.783	-32.66	48.99
Big	0.900	5.135	0.068	6.503	-27.10	33.40
Panel D : The 10 benchmark portfolios based on Book-to-market						
	Mean	Stdev.	Skewness	Kurtosis	Min	Max
Low	0.872	5.779	0.014	5.038	-27.98	38.77
2	0.978	5.543	-0.069	5.040	-26.83	34.82
3	0.981	5.396	-0.236	4.888	-27.43	31.23
4	0.957	6.092	1.254	15.822	-24.41	56.5
5	1.051	5.663	0.947	13.101	-28.27	46.63
6	1.095	6.240	0.914	16.961	-34.58	58.35
7	1.114	6.739	1.845	20.724	-33.65	61.68
8	1.264	7.011	2.261	25.470	-31.37	71.81
9	1.299	7.696	1.317	14.420	-39.05	63.95
High	1.402	9.384	2.393	24.293	-45.46	100.00

Table I, cont.						
Panel E : The 10 benchmark portfolios based on Momentum						
	Mean	Stdev.	Skewness	Kurtosis	Min	Max
1	0.323	9.627	1.912	18.201	-42.18	93.26
2	0.713	8.178	1.910	21.621	-40.61	80.08
3	0.729	7.043	1.571	20.017	-34.84	66.16
4	0.856	6.499	1.593	18.294	-32.23	62.66
5	0.860	6.028	1.275	17.976	-31.42	60.37
6	0.943	5.905	0.777	12.324	-31.01	45.04
7	1.025	5.651	0.131	7.672	-34.67	37.03
8	1.161	5.439	0.047	4.783	-27.10	32.48
9	1.267	5.749	-0.299	3.731	-26.91	32.90
10	1.599	6.562	-0.502	2.311	-28.64	28.58
Panel F : The 10 benchmark portfolios based on long-term reversal						
	Mean	Stdev.	Skewness	Kurtosis	Min	Max
1	1.498	8.806	2.321	21.258	-40.64	91.95
2	1.309	7.944	2.981	31.041	-35.18	88.39
3	1.279	6.964	1.961	21.698	-37.35	64.64
4	1.090	6.184	1.672	20.875	-33.67	61.11
5	1.150	6.222	2.029	23.157	-28.53	64.36
6	1.045	5.761	1.016	14.630	-28.59	49.56
7	1.064	5.858	1.419	17.701	-27.43	60.44
8	1.049	5.779	0.745	10.077	-30.37	41.62
9	0.932	5.838	0.210	6.728	-29.77	37.23
10	0.894	6.374	-0.350	3.745	-34.16	30.78

As in the second part of the simulation study, we now do not assume a normal distribution but rather analyze the empirical return distribution, which clearly is non-normal (see the skewness and kurtosis statistics in Table I). Also, we now use the value-weighted market index rather than the EP, TP or LP test portfolios constructed for use in the simulations.

4.3 Beta sorted portfolios

In this section, we analyze if the CRSP all-share index (a value-weighted average of all common stocks listed on the NYSE, AMEX, and NASDAQ markets) is efficient relative to the ten beta-sorted portfolios (which were also used for the simulations in Section 2.5). The set of beta-sorted portfolios is the first of 6 datasets that will be analysed, and will be discussed in somewhat greater detail than the others, in order to keep maintain an acceptable length.

Beta-sorted portfolios have been used extensively to test the Sharpe-Lintner-Mossin Capital Asset Pricing Model (CAPM); see Black, Jensen and Scholes (1972), Friend and Blume (1973), Fama and MacBeth (1973), Reinganum (1981) and Fama and French (1992), among others. The empirical results suggest that the CAPM is violated, because the spread in the means is too small relative to the spread in the betas. In other words, by buying low-beta stocks and selling high-beta stocks, we can "beat the market" (achieve a higher Sharpe-ratio than the market portfolio).

While portfolios formed on market capitalization and book-to-market equity ratio are more popular in research, there are good reasons to start elsewhere: even without taking a firm position on the existing empirical evidence regarding the associated anomalies, size and book-to-market portfolios generally lack variation in their betas and hence involve a high probability of Type I error. Furthermore, they exhibit a strong cyclical pattern in their risk profiles (see for example Lettau and Ludvigson, 2001), and a data snooping bias may arise when sorting stocks on characteristics that are known to be correlated with average returns (see for example Lo and MacKinlay, 1990, though in many cases such a bias may be unavoidable, such as with the momentum portfolios analysed below), and the portfolios assign a high weight to small cap stocks that represent only a small fraction of the total market capitalization (see for instance Loughran, 1997).

We apply the (unrestricted) SSD, TSD and GRS tests to the full sample from January 1933 to December 2002 (840 months), as well as to four non-overlapping subsamples of 210 months (January 1933 – June 1950, July 1950 - December 1967, January 1968 – June 1985, and July 1985 - December 2002). Table II shows the resulting pricing errors and the p-values.

Table II

Pricing Errors and p-values

The table shows the pricing errors and p-values for the SSD, TSD and GRS tests for efficiency of the CRSP all-share index relative to the ten beta portfolios in the full sample and four non-overlapping subsamples of 210 months. The SSD and TSD kernels are based one iteration, while the p-values are based on the resulting weighting matrix.

Panel A: SSD Efficiency						
		Jan 1933-Dec 2002	Jan 1933-Jun 1950	Jul 1950-Dec 1967	Jan 1968-Jun 1985	Jul 1985-Dec 2002
Pricing errors	Low	0.046	0.066	0.072	0.157	0.005
	2	0.043	-0.188	0.009	0.031	0.222
	3	-0.015	-0.286	0.005	-0.122	0.141
	4	-0.138	-0.392	-0.051	-0.192	0.046
	5	-0.038	-0.123	-0.030	-0.293	0.208
	6	-0.144	-0.228	-0.041	-0.386	0.054
	7	-0.250	-0.289	-0.021	-0.567	-0.065
	8	-0.356	-0.420	-0.309	-0.602	-0.216
	9	-0.434	-0.501	-0.314	-0.725	-0.252
	High	-0.552	-0.455	-0.417	-0.760	-0.776
	p-value	0.129	0.237	0.424	0.257	0.533

Panel B: TSD Efficiency						
		Jan 1933-Dec 2002	Jan 1933-Jun 1950	Jul 1950-Dec 1967	Jan 1968-Jun 1985	Jul 1985-Dec 2002
Pricing errors	Low	0.075	0.128	0.066	0.203	0.033
	2	0.109	-0.064	0.006	0.138	0.290
	3	0.050	-0.106	0.083	-0.035	0.219
	4	-0.087	-0.230	-0.025	-0.105	0.136
	5	0.083	0.155	0.124	-0.220	0.326
	6	-0.017	0.043	0.057	-0.340	0.152
	7	-0.125	0.011	0.059	-0.514	0.033
	8	-0.157	-0.033	-0.244	-0.559	-0.127
	9	-0.244	-0.105	-0.204	-0.698	-0.182
	High	-0.343	-0.036	-0.374	-0.738	-0.704
	p-value	0.057	0.224	0.181	0.162	0.376

Panel C: GRS (MV Efficiency)						
		Jan 1933-Dec 2002	Jan 1933-Jun 1950	Jul 1950-Dec 1967	Jan 1968-Jun 1985	Jul 1985-Dec 2002
Pricing errors	Low	0.249	0.342	0.293	0.391	0.019
	2	0.191	0.053	0.181	0.295	0.277
	3	0.149	0.008	0.293	0.121	0.205
	4	0.004	-0.152	0.023	0.044	0.119
	5	0.144	0.136	0.216	-0.051	0.306
	6	-0.001	-0.013	0.138	-0.179	0.128
	7	-0.096	-0.099	0.091	-0.314	0.005
	8	-0.253	-0.250	-0.121	-0.350	-0.161
	9	-0.285	-0.269	-0.057	-0.456	-0.219
	High	-0.475	-0.242	-0.241	-0.449	-0.749
	p-value	0.003	0.064	0.018	0.097	0.441

In line with the existing evidence for beta-sorted portfolios, we must reject the CAPM in the full sample, with a GRS p-value of 0.3%. As expected, the low-beta stocks are substantially underpriced (or have a positive pricing error) and the high-beta stocks are substantially overpriced (or have a negative pricing error).

Interestingly, the evidence against SSD and TSD efficiency of the market portfolio is substantially weaker, with p-values of 12.9% and 5.7% respectively, which is within the range of conventional significance levels. Note that the SSD pricing errors for the high-beta portfolios actually worsen relative to the GRS results. This is possible because the high-beta portfolios are more volatile and more strongly correlated than the low-beta portfolios and hence their pricing errors are assigned a low weight relative to those of the low-beta portfolios. Thus, the improvement of the errors for the low-beta portfolios outweighs the deterioration of the negative errors for the high-beta portfolios. By contrast, using the TSD criterion, improvements occur both for the low-beta portfolios and the high-beta portfolios.

Figure 2 shows the pricing kernels or marginal utility functions associated with the three efficiency tests in the full sample. The TSD pricing kernel is a three-piece linear function that kinks at returns of about -13% and –23%.[4] The SSD pricing kernel is very similar, apart from having a stepwise shape that deviates from convexity (hence violating skewness preference). The large increments of the SSD and TSD kernels for large negative returns suggests that part of the GRS pricing errors can be explained by a high aversion for tail risk. These findings confirm the pattern in Table 1: the low-beta portfolios, which appear underpriced in mean-variance terms, have a negatively skewed return distribution, while the high-beta portfolios, which appear overpriced in mean-variance terms, have a high positive skewness. For example, the lowest beta-decile portfolio has a skewness of –0.754, while the highest-beta portfolio with the highest-beta stocks has a skewness of 0.814.

The regular beta systematically underestimates tail risk for low-beta stocks and overestimates tail risk for high-beta stocks. For example, the lowest-beta portfolio has

[4] It should be noted that this kernel is not differentiable due to the kinks. However, as discussed in Chapter 2, differentiability is not critical to our analysis. As explained in Section 2.2C, the gradient vector $\nabla \boldsymbol{u}$ is the only aspect of the kernel that is actually used in the analysis. We may smoothen the kernel in the neighbourhood of the kinks without affecting this gradient vector. The same is true for the SSD kernel.

a regular beta of 0.597 and a tail beta of 0.858, while the highest-beta portfolio has a regular beta of 1.763 and a tail beta of 1.583. Similar conclusions can be drawn for the second-highest-beta and second-lowest-beta portfolios, and so on. The SSD and TSD tests pick up this pattern of tail behavior, while the GRS test considers mean and variance only. Figure 3 illustrates this in a different way: this mean-beta plot shows the theoretical relation as predicted by the CAPM as a solid line, and contains two sets of points, one based on the regular beta of the portfolios (the open squares), and one on the tail betas (the solid triangles). The actual Security Market Lines are omitted for the sake of clarity. We see that in 9 out of ten cases the tail beta is the risk measure that yields a result closer to a linear risk-return relationship than the regular beta.

The results for the subsamples confirm the findings for the full sample. In every subsample, the same pattern emerges: low-beta stocks are underpriced in mean-variance terms and the high-beta stocks are overpriced. In three of the four subsamples, the market portfolio is significantly mean-variance inefficient. By contrast, the market becomes efficient if we use the stochastic dominance criteria. The SSD and TSD criteria are especially successful at reducing the pricing errors of the low-beta stocks. Apparently, the risk of these stocks is higher than suggested by their market beta. Indeed, a follow-up analysis (not reported here) shows that the low-beta stocks have relatively high tail betas in all subsamples.

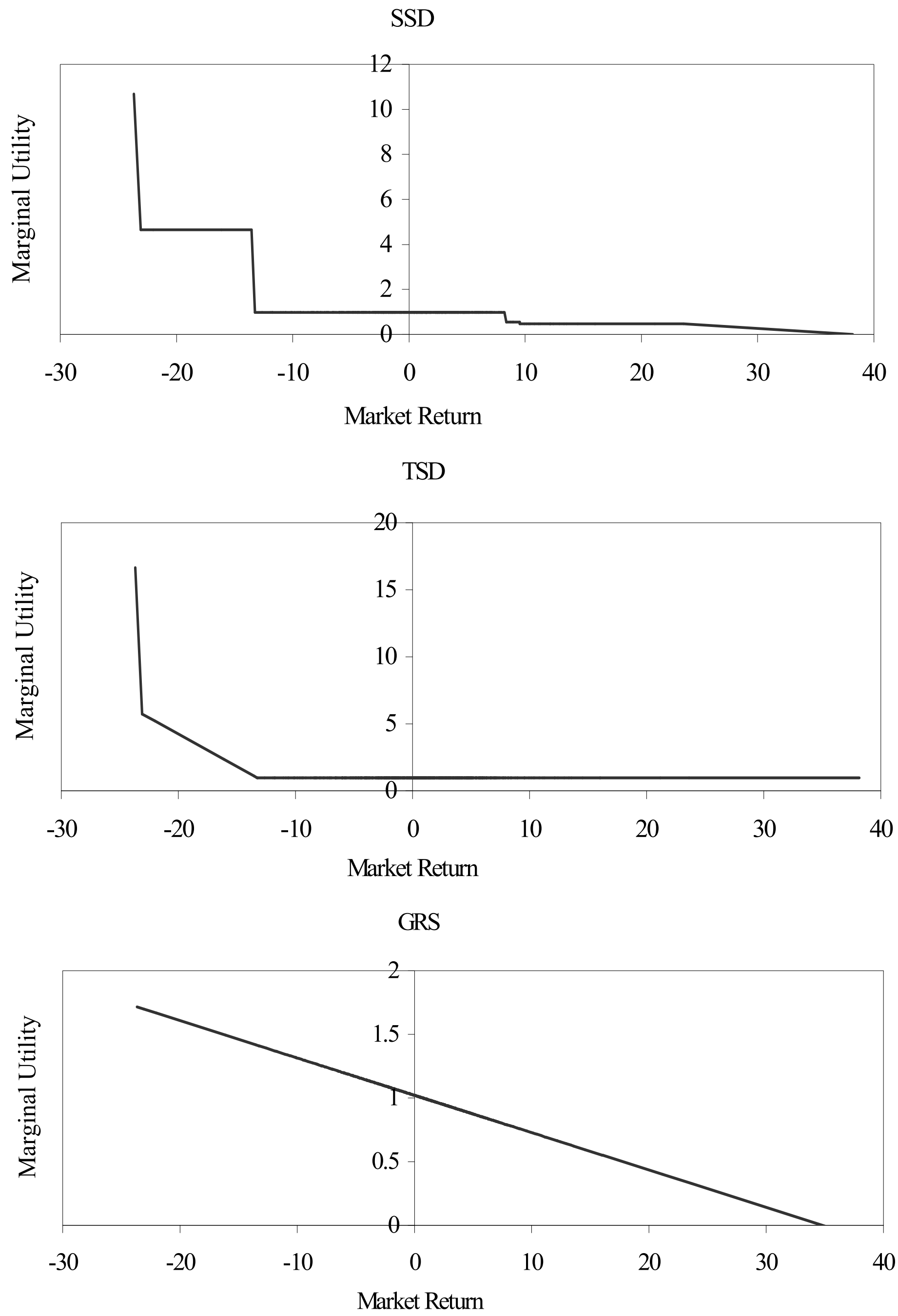


Figure 2

Pricing Kernels

The figure shows the pricing kernels or marginal utility functions for the SSD (top panel), TSD (middle panel) and GRS (lower panel) tests in the sample of beta portfolios from January 1933 to December 2002 (840 months). For the SD tests, the marginal utility functions are obtained by means of linear interpolation from the optimal gradient vectors. For the GRS test, the marginal utility function is the linear function $u'(x) = a + bx$, with a and b found by solving $T^{-1}\nabla \boldsymbol{u}^{\mathrm{T}} \mathbf{1}_T = 1$ and $T^{-1}\nabla \boldsymbol{u}^{\mathrm{T}} \mathbf{X}^{\mathrm{T}} \boldsymbol{\tau} = 0$.

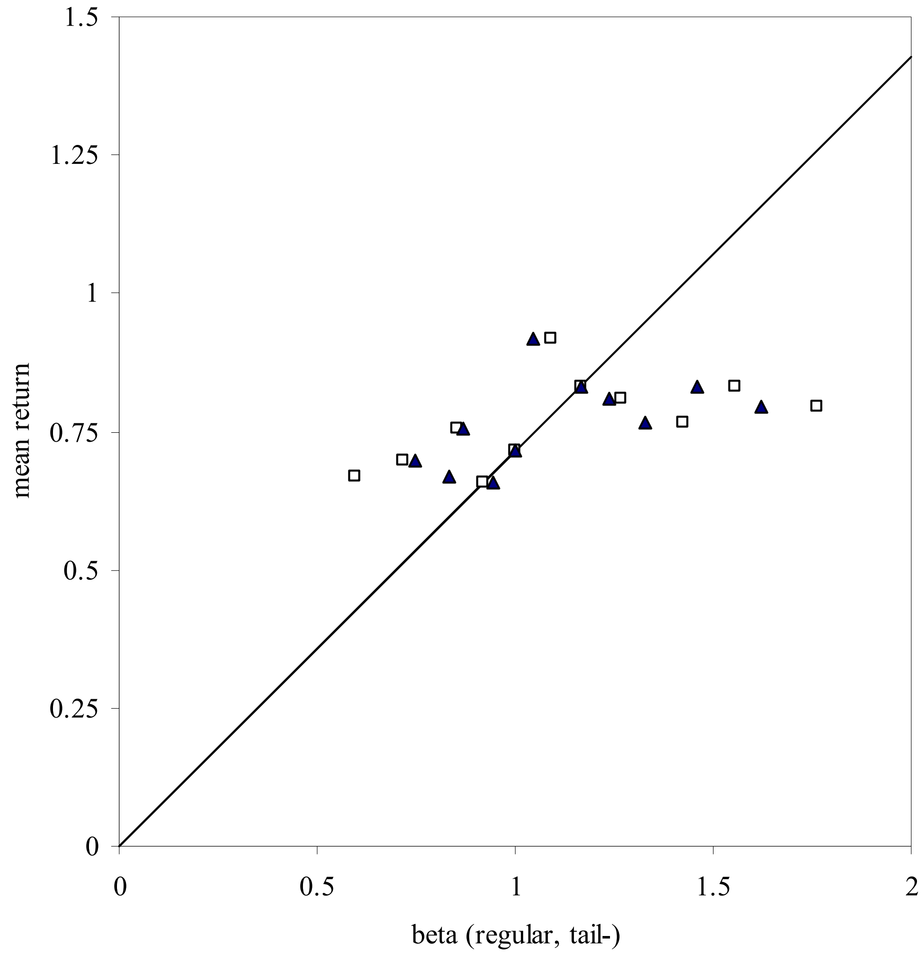


Figure 3

Tail Betas vs. Regular Betas

The figure shows the mean returns plotted against the tail betas (triangles) en regular betas (squares) for the beta-sorted portfolios for the sample period from January 1933 to December 2002 (840 months). The tail beta is measured as the co-lower partial moment (see equation 16) with a target rate of return *m* of –10%. A line depicting the linear risk-return relationship (based on the market return) is added for ease of interpretation.

As for the TFS results, which can be found in Table III, we see that in the whole sample, SSD-TFS is rejected (though the p-value is 3.7%, and the sub-periods indicate that only the last period is responsible for this rejection). Hence we can conclude that some representative investor models are supported by the data (efficiency), but not that all models would yield that result (TFS). We are therefore facing a rather large uncertainty with regard to the specification; looking at the TSD tests we may be able to narrow the set of admissible models.

However, the TSD tests offer a mirror image of the SSD results. TFS is rejected, and again the years 1985-2002 seem to be the culprits. Efficiency is not rejected (5.7%, it is only logical that the p-value is lower for the more stringent TSD specification, which allows less liberty to the kernel). Our conclusion has to be that while a non-satiable, risk-averse and skewness-averse can explain the beta-sorted portfolios, these characteristics are not restrictive enough; some of the models that fall in this category will offer a risk-based explanation of the market returns, yet others won't. In figure 1, this corresponds to scenario IV.

In summary, our analysis suggests that an TSD compatible representative investor model can explain why the stock market portfolio is mean-variance inefficient relative to beta portfolios, and the explanatory power of the model appears to come form tail risk that remains uncaptured by variance. The relatively small spread in the means relative to the spread in the betas is consistent with the tail betas of low-beta stocks being higher than their regular betas and the tail betas of high-beta stocks being lower than their regular betas. This finding implies that a straightforward generalization of the CAPM that accounts for tail risk may suffice to describe the cross-section of stock returns. However, one must thread carefully, as the remains a notable uncertainty about the specification; the TFS results show that not all models in this category will do, and that more restrictions (or parametrisation) may be required.

Table III

Pricing Errors and p-values

The table shows the pricing errors and p-values for the SSD-TFS, and TSD-TFS tests, using the CRSP all-share index relative to the ten beta-sorted portfolios in the full sample and four non-overlapping subsamples.

Panel A: SSD-TFS						
		beta sorted portfolios				
		Jan 1933-Dec 2002	Jan 1933-Jun 1950	Jul 1950-Dec 1967	Jan 1968-Jun 1985	Jul 1985-Dec 2002
Pricing errors	Small	0.242	0.160	0.335	0.377	0.060
	2	0.192	-0.043	0.190	0.259	0.281
	3	0.146	-0.067	0.292	0.106	0.209
	4	0.019	-0.184	0.000	0.041	0.114
	5	0.160	0.213	0.209	-0.047	0.279
	6	0.022	0.104	0.119	-0.145	0.121
	7	-0.097	0.044	0.064	-0.294	-0.003
	8	-0.261	0.043	-0.114	-0.310	-0.196
	9	-0.284	-0.021	-0.028	-0.405	-0.222
	Large	-0.467	0.059	-0.202	-0.377	-0.839
p-value		0.037	0.920	0.259	0.120	0.000
Panel B: TSD-TFS						
		beta sorted portfolios				
		Jan 1933-Dec 2002	Jan 1933-Jun 1950	Jul 1950-Dec 1967	Jan 1968-Jun 1985	Jul 1985-Dec 2002
Pricing errors	Small	0.246	0.132	0.293	0.375	0.019
	2	0.190	-0.049	0.181	0.287	0.277
	3	0.152	-0.080	0.293	0.121	0.205
	4	0.004	-0.198	0.023	0.051	0.119
	5	0.149	0.218	0.216	-0.048	0.306
	6	0.000	0.097	0.138	-0.173	0.128
	7	-0.094	0.073	0.091	-0.317	0.005
	8	-0.251	0.055	-0.121	-0.351	-0.161
	9	-0.284	0.002	-0.057	-0.455	-0.219
	Large	-0.469	0.069	-0.241	-0.450	-0.749
p-value		0.019	0.928	0.369	0.175	0.007

4.4 Size and book-to-market sorted portfolios

Although not our first choice, it would be an omission not to devote a section on the analysis of size- (market equity) and book-to-market-(value) sorted portfolios. Again we investigate if the CRSP all-share index is efficient relative to ten size-sorted portfolios, and relative to 10 value-sorted portfolios, and then turn to the analysis of the existence of Two-Fund Separation.

MV inefficiency of size and value-sorted portfolios has been documented by, among others, Basu (1977), Banz (1981) and Fama and French (1992) and (1993), as well as a multitude of papers derived from the ones cited above. However, the inefficiency also depends crucially on the period under investigation. As a matter of fact, in our dataset, the complete sample is not MV-inefficient, and the more anomalous period of the 1960s and 70s is divided between 2 subperiods, though the subperiod that includes the 70s and part of the 60s has a markedly lower p-value (12.5%, at least 50% below the rest). Splitting the periods differently, or excluding the first 6 years, would result in inefficiency. Apart from this problem, section 4.3 noted some objections to the use of these portfolios[5], yet the analysis of this data remains popular. Our results are in line with those of earlier studies (see for instance Gibbons, Ross and Shanken (1989)), and we expect ex ante to find a contrast between these size-sorted portfolios and their cousins, the value-sorted portfolios. (which are harder to explain; see the references above). When we replace the asset span with the value-sorted portfolios we see that this effect is also somewhat period-dependent, but the troublesome periods are in different decades. Still, it seems useful to apply our efficiency tests if only because of the basis on first principles (see Chapter 1), as well as the TFS test, because the latter test can determine the support for alternative representative investor models—models that cannot be dismissed at forehand due to the uncertainty surrounding the appropriate model specification.

[5] One could also add that the economic interpretation is sometimes problematic (why would 'book-to-market' have a real effect?), but this objection holds for several anomalies, and does not negate the fact that a model like the CAPM aspires to explain all portfolios, regardless of the way they are formed.

Table IV

Pricing Errors and p-values

The table shows the pricing errors and p-values for the SSD, TSD and GRS tests for efficiency of the CRSP all-share index relative to the ten size portfolios in the full sample (956 months) and four non-overlapping subsamples.

Panel A: SSD Efficiency

		Size sorted portfolios					Book-to-market sorted portfolios				
		Jan 1927-Dec 2002	Jan 1927-Dec 1945	Jan 1946-Dec 1964	Jan 1965-Dec 1983	Jan 1984-Dec 2002	Jan 1927-Dec 2002	Jan 1927-Dec 1945	Jan 1946-Dec 1964	Jan 1965-Dec 1983	Jan 1984-Dec 2002
Pricing errors	Small	0.171	0.246	-0.071	-0.039	-0.054	0.236	0.348	0.900	0.257	0.330
	2	0.069	-0.088	-0.092	-0.161	-0.074	0.359	0.479	0.855	0.088	0.470
	3	0.073	-0.103	-0.067	-0.132	-0.119	0.348	0.446	0.905	0.185	0.532
	4	0.073	-0.035	-0.078	-0.099	-0.193	0.315	0.471	0.811	0.147	0.494
	5	0.055	-0.125	-0.085	-0.169	-0.081	0.378	0.535	1.086	0.139	0.543
	6	0.046	-0.036	-0.048	-0.215	-0.164	0.493	0.618	0.927	0.212	0.529
	7	0.053	-0.062	-0.057	-0.253	-0.049	0.451	0.687	0.723	0.152	0.752
	8	0.037	-0.151	-0.032	-0.254	-0.096	0.633	0.973	1.176	0.234	0.624
	9	0.010	-0.159	-0.003	-0.305	-0.037	0.577	1.067	0.826	0.257	0.674
	Large	-0.007	-0.163	0.006	-0.423	-0.076	0.734	1.343	0.995	0.344	0.702
p-value		0.997	0.980	1.000	0.647	0.750	0.781	0.969	0.004	0.969	0.626

Panel B: TSD Efficiency

		Size sorted portfolios					Book-to-market sorted portfolios				
		Jan 1927-Dec 2002	Jan 1927-Dec 1945	Jan 1946-Dec 1964	Jan 1965-Dec 1983	Jan 1984-Dec 2002	Jan 1927-Dec 2002	Jan 1927-Dec 1945	Jan 1946-Dec 1964	Jan 1965-Dec 1983	Jan 1984-Dec 2002
Pricing errors	Small	0.225	1.036	-0.150	0.026	-0.159	0.517	0.608	0.987	0.177	0.513
	2	0.094	0.525	-0.164	-0.094	-0.122	0.627	0.810	0.942	0.052	0.697
	3	0.089	0.367	-0.114	-0.068	-0.143	0.602	0.743	0.931	0.193	0.721
	4	0.089	0.374	-0.120	-0.040	-0.199	0.546	0.692	0.861	0.125	0.719
	5	0.075	0.261	-0.142	-0.118	-0.075	0.663	0.700	1.262	0.141	0.730
	6	0.056	0.315	-0.084	-0.161	-0.146	0.699	0.825	1.096	0.285	0.679
	7	0.065	0.220	-0.090	-0.206	0.014	0.696	0.901	0.917	0.166	0.854
	8	0.035	0.087	-0.039	-0.212	-0.020	0.875	1.145	1.325	0.340	0.722
	9	0.007	0.045	0.004	-0.264	0.050	0.883	1.307	1.118	0.334	0.796
	Large	-0.023	-0.054	0.027	-0.407	0.025	0.979	1.501	1.122	0.438	0.902
p-value		0.978	0.867	0.968	0.643	0.575	0.003	0.838	0.000	0.827	0.279

Panel C: GRS (MV Efficiency)

		Size sorted portfolios					Book-to-market sorted portfolios				
		Jan 1927-Dec 2002	Jan 1927-Dec 1945	Jan 1946-Dec 1964	Jan 1965-Dec 1983	Jan 1984-Dec 2002	Jan 1927-Dec 2002	Jan 1927-Dec 1945	Jan 1946-Dec 1964	Jan 1965-Dec 1983	Jan 1984-Dec 2002
Pricing errors	Small	0.268	1.086	-0.134	0.657	-0.084	0.587	0.669	1.081	0.149	0.503
	2	0.134	0.571	-0.150	0.487	-0.034	0.691	0.867	1.025	0.215	0.707
	3	0.127	0.408	-0.099	0.495	-0.057	0.684	0.800	1.046	0.235	0.750
	4	0.125	0.412	-0.106	0.503	-0.113	0.670	0.774	0.946	0.256	0.733
	5	0.111	0.299	-0.128	0.452	0.011	0.760	0.767	1.290	0.243	0.738
	6	0.091	0.352	-0.070	0.303	-0.065	0.817	0.907	1.140	0.513	0.699
	7	0.099	0.254	-0.075	0.232	0.095	0.841	1.000	0.930	0.509	0.893
	8	0.067	0.119	-0.026	0.184	0.063	1.015	1.245	1.373	0.670	0.743
	9	0.038	0.077	0.017	0.018	0.126	1.023	1.412	1.077	0.703	0.810
	Large	0.004	-0.028	0.041	-0.142	0.099	1.149	1.607	1.114	0.860	0.917
p-value		0.977	0.847	0.970	0.125	0.418	0.124	0.993	0.236	0.783	0.829

We apply the SSD, TSD and GRS efficiency tests to the full sample from January 1927 to December 2002 (912 months), as well as to four non-overlapping subsamples of 228 months: January 1927 – December 1945, January 1946 – December 1964, January 1965 – December 1983, and January 1984 – December 2002. Table IV shows the resulting pricing errors and p-values for the three tests.[6] Table V shows the results for the SSD-TFS and TSD-TFS tests.

Table V

Pricing Errors and p-values

The table shows the pricing errors and p-values for the SSD-TFS, and TSD-TFS tests, using the CRSP all-share index relative to the ten size portfolios in the full sample and four non-overlapping subsamples, as well as the Book-to-market sorted portfolios.

Panel A: SSD - TFS											
		Size sorted portfolios					Book-to-market sorted portfolios				
		Jan 1927-Dec 2002	Jan 1927-Dec 1945	Jan 1946-Dec 1964	Jan 1965-Dec 1983	Jan 1984-Dec 2002	Jan 1927-Dec 2002	Jan 1927-Dec 1945	Jan 1946-Dec 1964	Jan 1965-Dec 1983	Jan 1984-Dec 2002
Pricing errors	Small	0.210	0.801	-0.125	0.705	0.070	0.690	0.816	1.059	0.216	0.597
	2	0.084	0.374	-0.134	0.513	0.084	0.794	1.027	1.001	0.272	0.764
	3	0.103	0.268	-0.083	0.527	0.042	0.778	0.957	0.996	0.282	0.819
	4	0.111	0.296	-0.090	0.541	-0.039	0.774	0.977	0.914	0.310	0.775
	5	0.101	0.195	-0.110	0.479	0.078	0.854	0.931	1.249	0.287	0.765
	6	0.079	0.267	-0.061	0.325	-0.013	0.928	1.101	1.091	0.560	0.761
	7	0.086	0.193	-0.059	0.233	0.105	0.965	1.248	0.857	0.562	0.952
	8	0.058	0.069	-0.018	0.195	0.060	1.121	1.481	1.327	0.709	0.780
	9	0.031	0.046	0.025	0.013	0.122	1.155	1.690	1.027	0.752	0.868
	Large	-0.002	-0.018	0.035	-0.156	0.079	1.257	1.878	1.052	0.883	0.984
p-value		0.685	0.538	0.627	0.001	0.807	0.000	0.000	0.000	0.000	0.000
Panel B: TSD - TFS											
		Size sorted portfolios					Book-to-market sorted portfolios				
		Jan 1927-Dec 2002	Jan 1927-Dec 1945	Jan 1946-Dec 1964	Jan 1965-Dec 1983	Jan 1984-Dec 2002	Jan 1927-Dec 2002	Jan 1927-Dec 1945	Jan 1946-Dec 1964	Jan 1965-Dec 1983	Jan 1984-Dec 2002
Pricing errors	Small	0.268	1.086	-0.134	0.605	-0.029	0.587	0.669	1.067	0.149	0.503
	2	0.134	0.571	-0.150	0.452	0.001	0.691	0.867	1.016	0.215	0.707
	3	0.127	0.408	-0.099	0.463	-0.030	0.684	0.800	1.045	0.235	0.750
	4	0.125	0.412	-0.106	0.473	-0.091	0.670	0.774	0.942	0.256	0.733
	5	0.111	0.299	-0.128	0.418	0.031	0.760	0.767	1.288	0.243	0.738
	6	0.091	0.352	-0.070	0.286	-0.053	0.817	0.907	1.143	0.513	0.699
	7	0.099	0.254	-0.075	0.223	0.098	0.841	1.000	0.933	0.509	0.893
	8	0.067	0.119	-0.026	0.184	0.061	1.015	1.245	1.374	0.670	0.743
	9	0.038	0.077	0.017	0.037	0.117	1.023	1.412	1.086	0.703	0.810
	Large	0.004	-0.028	0.041	-0.129	0.093	1.149	1.607	1.152	0.860	0.917
p-value		0.752	0.520	0.561	0.048	0.829	0.000	0.000	0.000	0.000	0.000

[6] The SSD and TSD kernels are based one iteration, while the p-values are based on the resulting weighting matrix.

For size portfolios in the full sample, we see results conforming to scenario I (see figure 1): SSD-TFS cannot be rejected - the p-value is 68,5%, well above conventional levels, so all representative investor models which include risk-aversion and non-satiation are valid. These portfolios make the task of explaining the market portfolio relatively easy; a large collection of specifications will do the job. Logically, it should follow that there should be SSD efficiency (if all SSD models are possible, there will be of course an SSD model that makes the market efficient; p-value 99,7%), but also TSD-TFS (p-value 75,2%), as the TSD specifications are a subset of the SSD ones, from which TSD efficiency follows in turn with a p-value of 97,8%. These results are confirmed in each subperiod, with the exception of the 1965-1983 period, in which TFS is rejected (though only just for the TSD variant), but efficiency isn't. This corresponds to scenario IV; for a more detailed description of this situation, see section 4.2. We can offer no explanation why this subperiod gives such different results, nor is this period problematic in other datasets.

One should note that the good fits in the complete sample are obtained despite sizeable pricing errors for the small cap, high-beta portfolios (also shown in Figure 1). The returns of these portfolios are highly volatile and highly correlated, and hence the pricing errors of these portfolios are assigned a relatively low weight.

At this point it may be interesting to compare the kernels for the GRS efficiency test with those from the SSD efficiency and TFS tests, which are shown in Figure 2. The mean-variance kernel is linear, with the intercept and slope reflecting the level of the sample equity premium. Apart from the stepwise shape, the SSD kernel comes reasonably close to the mean-variance kernel. As both kernels are designed to minimize the pricing errors, this is not surprising. By contrast, the TFS kernel has a very different shape, with risk neutrality for losses and strong risk aversion for gains. This kernel severely violates decreasing absolute risk aversion (DARA), and hence we may question if it is plausible. However, the definition of TFS does not require DARA and thus admits such an "exotic" kernel. Actually, it is very encouraging that we cannot reject market portfolio efficiency even for such a kernel; all other admissible kernels yield a similar or better fit and imposing DARA can only further improve the fit. Still, further research may focus on modified definitions of TFS that include preference assumptions in addition to nonsatiation and risk aversion (also see the concluding remarks).

Given the seemingly poor fit for the mean-tail risk model in Figure 4, it seems surprising that the TFS test does not identify a mean-tail risk model as giving the worst possible fit. This result can be explained by the high uncertainty surrounding the pricing errors for mean-tail risk models. Accounting for this uncertainty, the market portfolio is not significantly inefficient for mean-tail risk investors, despite the large pricing errors.[7] In fact, the TFS model performs worse than the mean-tail risk model, despite its smaller—but more reliable—pricing errors.

The inability to reject TFS implies that the evidence in favor of the mean-variance model carries over to every representative investor model. This finding is comforting for proponents of the CAPM given the uncertainty surrounding the appropriate model specification. For example, measures of downside risk or tail risk seem more plausible risk measures than variance. Our results suggest that, like the mean-variance model, models that use such alternative risk measures cannot be rejected and the mean-variance model can in fact serve as an acceptable proxy for such models. In this respect, our findings confirm the conclusion by Levy and Markowitz (1979) that the mean-variance rule generally is a good substitute for the expected utility rule for typical asset return data.[8]

[7] We stress again that it is important to account for the shape of the kernel when correcting for the uncertainty of the errors (see Section 3.2B). In other words, the weighting of the errors is not only directly dependent on the data, but also indirectly through the kernel $\boldsymbol{\beta}$. To demonstrate this point, if we use $\hat{\boldsymbol{\Omega}}(\boldsymbol{\beta})$ to weight the errors of the TFS test applied to the size dataset, the p-value is 0.675 and we cannot reject TFS. By contrast, if we replace $\hat{\boldsymbol{\Omega}}(\boldsymbol{\beta})$ with $\hat{\boldsymbol{\Omega}}(\mathbf{1}_N)$, which boils down to ignoring the shape of the kernel, the p-value drops below 0.01 and the model must be rejected. By contrast, our approximation $\phi(\boldsymbol{\beta})\hat{\boldsymbol{\Omega}}(\mathbf{1}_T)$, which partially corrects for the shape of the kernel, gives a much better approximation with a p-value of 0.605. The pattern described here is similar to that found in other setups, but we must leave a complete analysis of our approximation for future research.

[8] Our results are even stronger than those of Levy and Markowitz, because their conclusions did not apply for 'extreme' utility functions such as $u(x) - e^{-10(1+x)}$, while our results apply for all monotone and concave utility functions. Also, our TFS test allows for leverage and diversification, which increases the range of feasible return distributions beyond that considered by Levy and Markowitz

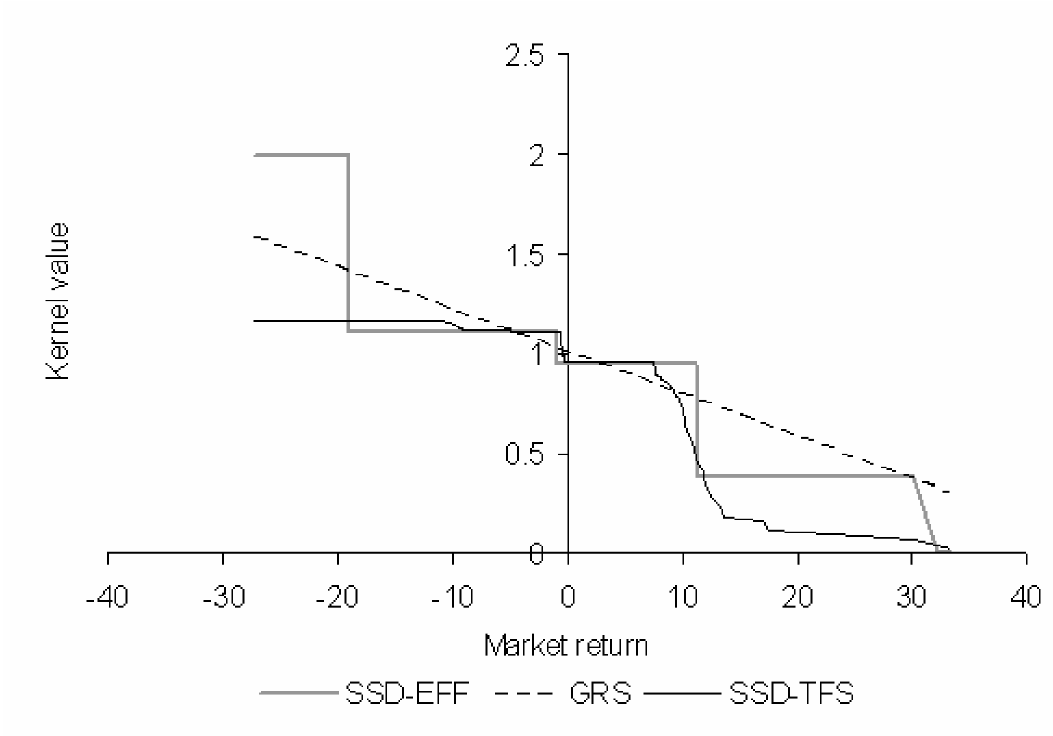


Figure 4

Pricing Kernels

The figure shows the pricing kernels or marginal utility functions for the SSD test, GRS test and TFS test in the sample of size portfolios from January 1927 to December 2002 (912 months). For the SSD and TFS tests, the marginal utility functions are obtained by means of linear interpolation from the optimal gradient vectors. For the GRS test, the marginal utility function is the linear function $u'(x) = a + bx$, with a and b found by solving $T^{-1}\nabla u^{\mathrm{T}}\boldsymbol{e} = 1$ and $T^{-1}\nabla u^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\boldsymbol{\tau} = 0$.

The Book-to-market sorted portfolios (or value portfolios, are they are often called) show an altogether different picture. For the complete sample, SSD-TFS is completely rejected with a p-value of 0%, however, SSD-efficiency cannot be rejected (78,1%). The conclusion must be that the specification needed to explain the market vis-à-vis the value sorted portfolios will conform to the principles of risk-aversion and non-satiation, but that only a subset of these kernels will do. We can also exclude any kernel based on skweness-aversion, as not only TSD-TFS is rejected (again a p-value of 0) but TSD efficiency as well (0.3%). We therefore end up in scenario V, which reflects a large uncertainty regarding the robustness of the specifications; we only know that a part of the SSD-compatible kernels will explain the market, but we cannot draw any conclusions that narrow this set, except for the exclusion of TSD specifications.

The results in the subperiods are somewhat more encouraging, as TSD efficiency cannot be rejected for all subperiods, except for the years 1946-1964. The latter period seems problematic in several datasets, see below. TFS is still a bridge too far, but the ability to narrow the search to TSD-compatible specifications could be of help.

4.5 Momentum based portfolios

The momentum sorted portfolios are somewhat of an enigma in the field of Asset Pricing. The corresponding anomaly is described by Jigadeesh and Titman (1993); also see Griffin et al. (2003) in an international context. The point is that if one buys the stocks that have performed well over the past year, and short-sells the ones that didn't, excess returns can be obtained. Of course, whether a return qualifies as such depends on the measurement of risk, but so far there have been very few authors who have been able to explain this anomaly based on risk.[9] The MV-framework will not provide an answer, as can been seen in Table VI. Only in the pre-WW II period the market portfolio seems MV-efficient relative to momentum portfolios.

Unfortunately, SD doesn't fare much better, see Tables VII and VII. For the complete sample, all tests result in a firm rejection of the null. Not only SSD-TFS is rejected, but efficiency as well: no support for a risk-averse representative investor can be found in the data. If we switch to the TSD criterion, the results are just as unimpressive. This conforms to Scenario II in figure 1.

Looking at the subperiods, the early period (1927-1945) actually supports efficiency for both SSD and TSD (71.7 and 49.8% respectively), but not TFS. Some well-behaved representative investor models will do nicely in this period, but not all. However, the simpler MV-framework cannot be rejected either based on this data. The next 2 periods offer the same results as the complete sample: no risk based explanation

[9] However, these portfolios involve a lot of trading activity. Consequently, the costs of actually pursuing such a strategy seem to be very high, and may compensate all gains to be made. Jigadeesh and Titman (2001) revisit the issue and reject data snooping, but lend cautions support for more behavorial-orientent explanations. However, this question is outside the scope of this thesis.

is possible, at least not with a globally risk-averse representative investor. We may need a risk-seeking model, or, as noted before, perhaps an emphasis on transaction costs. The last period (1984-2002) is different. SSD-TFS cannot be rejected (74.3%), yet SSD efficiency is rejected. Logically, this is inconsistent, and the problem is compounded by the fact that the same conclusions must be drawn for the TSD tests. Here we find ourselves in the relatively rare situation that the difference between the two test methods causes problems. The efficiency tests are based on a sum of squared (pricing) errors, while the TFS tests are based on the highest possible pricing error. Secondly, the TFS test uses the worst possible utility function, and hence inflates the error, while the efficiency tests use the best possible utility function, thus deflating the (sum of squared) errors. The problem lies with the efficiency tests[10]: by their use of the best utility function they effectively maximize the statistical size (probability of rejection when the null is correct) which is probably what is happening here. Lastly, both tests use asymptotic bounds for the statistical inference, which may cloud the picture somewhat as well. In all, the result can be explained on methodological grounds, but offers little insights in the empirical questions. If we were to accept the non-rejection of SSD-TFS we would have to accept all representative investor models, which seems peculiar given the other periods. On the other hand, if the momentum anomaly is indeed explainable by transaction costs, it may be worthwhile to investigate the (alleged) reduction in trading costs for many market participants in recent years. Yet admittedly, we're venturing perilously close to the realm of speculation here.

[10] The TFS test maximizes the statistical power (probability of rejecting the null hypothesis when the null is incorrect), so we can normally be reasonably confident of a non-rejection like we have here.

Table VI

Pricing Errors and p-values

The table shows the pricing errors and p-values for the SSD, TSD and GRS tests for efficiency of the CRSP all-share index relative to the ten momentum portfolios in the full sample (912 months) and four non-overlapping subsamples.

Panel A: SSD Efficiency						
		momentum sorted portfolios				
		Jan 1927-Dec 2002	Jan 1927-Dec 1945	Jan 1946-Dec 1964	Jan 1965-Dec 1983	Jan 1984-Dec 2002
Pricing errors	Small	-1.331	-1.300	-0.960	-0.923	-1.643
	2	-0.738	-0.937	-0.614	-0.748	-0.507
	3	-0.589	-1.151	-0.309	-0.518	-0.235
	4	-0.408	-0.791	-0.289	-0.514	-0.060
	5	-0.354	-0.671	-0.021	-0.644	-0.172
	6	-0.285	-0.531	-0.074	-0.567	-0.094
	7	-0.185	-0.355	-0.060	-0.664	0.040
	8	-0.055	-0.245	-0.062	-0.584	0.218
	9	-0.029	-0.149	-0.045	-0.373	0.044
	Large	0.185	0.022	0.136	-0.198	0.218
p-value		0.000	0.717	0.000	0.005	0.000

Panel B: TSD Efficiency						
		momentum sorted portfolios				
		Jan 1927-Dec 2002	Jan 1927-Dec 1945	Jan 1946-Dec 1964	Jan 1965-Dec 1983	Jan 1984-Dec 2002
Pricing errors	Small	-1.291	-1.136	-0.907	-0.923	-1.528
	2	-0.712	-0.786	-0.560	-0.748	-0.355
	3	-0.565	-1.015	-0.261	-0.518	-0.121
	4	-0.367	-0.565	-0.177	-0.514	0.031
	5	-0.318	-0.492	0.090	-0.644	-0.079
	6	-0.250	-0.366	0.037	-0.567	-0.039
	7	-0.138	-0.164	0.075	-0.664	0.095
	8	0.017	-0.051	0.116	-0.584	0.304
	9	0.053	0.076	0.123	-0.373	0.108
	Large	0.308	0.262	0.321	-0.198	0.395
p-value		0.000	0.498	0.000	0.005	0.000

Panel C: GRS (MV Efficiency)						
		momentum sorted portfolios				
		Jan 1927-Dec 2002	Jan 1927-Dec 1945	Jan 1946-Dec 1964	Jan 1965-Dec 1983	Jan 1984-Dec 2002
Pricing errors	Small	-1.022	-0.740	-0.886	-0.743	-1.428
	2	-0.467	-0.404	-0.545	-0.341	-0.275
	3	-0.334	-0.652	-0.239	-0.127	-0.053
	4	-0.160	-0.238	-0.161	-0.146	0.095
	5	-0.097	-0.157	0.109	-0.182	-0.016
	6	-0.014	-0.029	0.061	0.006	0.025
	7	0.106	0.201	0.104	-0.008	0.156
	8	0.267	0.302	0.142	0.197	0.366
	9	0.343	0.490	0.148	0.496	0.175
	Large	0.657	0.697	0.360	0.909	0.481
p-value		0.000	0.206	0.000	0.000	0.000

Table VII

Pricing Errors and p-values

The table shows the pricing errors and p-values for the SSD-TFS, and TSD-TFS tests, using the CRSP all-share index relative to the ten momentum portfolios in the full sample and four non-overlapping subsamples.

Panel A: SSD-TFS						
		momentum sorted portfolios				
		Jan 1927-Dec 2002	Jan 1927-Dec 1945	Jan 1946-Dec 1964	Jan 1965-Dec 1983	Jan 1984-Dec 2002
Pricing errors	Small	-1.078	-1.036	-0.894	-0.727	-1.377
	2	-0.517	-0.625	-0.547	-0.345	-0.175
	3	-0.356	-0.790	-0.215	-0.147	0.063
	4	-0.204	-0.383	-0.183	-0.158	0.138
	5	-0.111	-0.260	0.077	-0.172	-0.007
	6	-0.023	-0.110	0.068	0.019	-0.057
	7	0.092	0.190	0.112	0.002	0.119
	8	0.261	0.329	0.138	0.197	0.351
	9	0.362	0.609	0.120	0.514	0.118
	Large	0.701	0.866	0.391	0.932	0.397
p-value		0.000	0.000	0.002	0.000	0.743
Panel B: TSD-TFS						
		momentum sorted portfolios				
		Jan 1927-Dec 2002	Jan 1927-Dec 1945	Jan 1946-Dec 1964	Jan 1965-Dec 1983	Jan 1984-Dec 2002
Pricing errors	Small	-1.022	-0.740	-0.886	-0.743	-1.428
	2	-0.467	-0.404	-0.545	-0.341	-0.275
	3	-0.334	-0.652	-0.239	-0.127	-0.053
	4	-0.160	-0.238	-0.161	-0.146	0.095
	5	-0.097	-0.157	0.109	-0.182	-0.016
	6	-0.014	-0.029	0.061	0.006	0.025
	7	0.106	0.201	0.104	-0.008	0.156
	8	0.267	0.302	0.142	0.197	0.366
	9	0.343	0.490	0.148	0.496	0.175
	Large	0.657	0.697	0.360	0.909	0.481
p-value		0.000	0.011	0.034	0.000	0.192

4.6 Reversal based portfolios

The long-term reversal ‘anomaly’ was first described by De Bondt & Thaler (1985). It involves the observation that the momentum-effect (short-term winners tend to outperform short-term losers) becomes reversed after a longer time-frame: stocks that underperformed in the past (usually measured as 3 to 5 years) outperform the stocks that were ‘winners’ during that period. In this sense the long-term reversal dataset is a nice contrast to the momentum data. Should the reversal portfolios be as difficult to understand as the momentum portfolios, a risk-based explanation of stock returns would be much harder to maintain. However, our data indicates that while the short-term momentum effects cannot be explained, the longer term reversal portfolios do not offer these difficulties. Of course, this raises the question which period to choose. The momentum data uses the past year, but another dataset (short-term reversal, from the same sources) confirms the momentum results when looking at only the past month as the sorting criteria. For reasons of brevity, these results are not reported here, but can be obtained from the author.

As customary, we apply the SSD, TSD and GRS efficiency tests to the full sample (which is somewhat smaller due to the 5 year lead time required; it runs from January 1931 to December 2002, which are 864 months), as well as to four non-overlapping subsamples of 216 months: January 1931 – December 1948, January 1949 – December 1966, January 1967 – December 1984, and January 1985 – December 2002. Table VIII shows the resulting pricing errors and p-values for the three tests.[11] Table IX shows the results for the SSD-TFS and TSD-TFS tests.

The reversal results offer relatively few difficulties. SSD-TFS is not rejected for the complete sample or the subperiods at the 5% level, though at the 10% level the period 1931-1948 almost leads to a rejection (10,5%), and 1967-1984 does lead to a rejection at this level (6,5%). SSD efficiency is not rejected either, and even the MV framework is capable of explaining this dataset (this seems contrary to De Bondt & Thaler's conclusions, though presumably this could be explained by differences in methodology, as De Bondt & Thaler use an approach based on cumulative abnormal returns). Logically, TSD-TFS and TSD efficiency follows, conforming to scenario I.

[11] The SSD and TSD kernels are based one iteration, while the p-values are based on the resulting weighting matrix.

Table VIII

Pricing Errors and p-values

The table shows the pricing errors and p-values for the SSD, TSD and GRS tests for efficiency of the CRSP all-share index relative to the ten long-term and short-term reversal portfolios in the full sample (864 and 912 months respectively) and four non-overlapping subsamples.

Panel A: SSD Efficiency						
		Long-term reversal portfolios				
		Jan 1931-Dec 2002	Jan 1931-Dec 2002	Jan 1931-Dec 2002	Jan 1931-Dec 2002	Jan 1931-Dec 2002
Pricing errors	Small	0.167	0.399	-0.113	0.173	-0.027
	2	0.046	-0.099	0.067	-0.082	0.157
	3	0.084	0.046	0.075	0.065	0.069
	4	0.019	-0.132	0.022	0.052	-0.061
	5	0.052	-0.016	0.017	-0.039	0.065
	6	0.005	-0.193	-0.017	-0.063	0.043
	7	0.016	-0.242	-0.026	-0.046	0.015
	8	0.011	-0.041	-0.116	-0.118	-0.014
	9	-0.084	-0.256	-0.112	-0.310	-0.159
	Large	-0.105	-0.305	-0.139	-0.576	-0.213
p-value		0.922	0.680	0.948	0.203	0.875
Panel B: TSD Efficiency						
		Long-term reversal portfolios				
		Jan 1931-Dec 2002	Jan 1931-Dec 2002	Jan 1931-Dec 2002	Jan 1931-Dec 2002	Jan 1931-Dec 2002
Pricing errors	Small	0.301	0.929	-0.076	0.521	0.035
	2	0.179	0.399	0.110	0.176	0.224
	3	0.219	0.424	0.152	0.276	0.110
	4	0.085	0.098	0.061	0.235	-0.026
	5	0.135	0.266	0.061	0.156	0.136
	6	0.069	-0.016	0.044	0.085	0.112
	7	0.073	-0.033	0.033	0.092	0.086
	8	0.060	0.163	-0.073	0.014	0.062
	9	-0.081	-0.126	-0.074	-0.174	-0.108
	Large	-0.172	-0.279	-0.105	-0.433	-0.096
p-value		0.294	0.373	0.848	0.215	0.749
Panel C: GRS (MV Efficiency)						
		Long-term reversal portfolios				
		Jan 1931-Dec 2002	Jan 1931-Dec 2002	Jan 1931-Dec 2002	Jan 1931-Dec 2002	Jan 1931-Dec 2002
Pricing errors	Small	0.412	0.842	-0.013	0.148	0.945
	2	0.217	0.342	0.029	0.173	0.776
	3	0.258	0.382	0.103	0.345	0.534
	4	0.104	0.090	0.103	0.204	0.265
	5	0.178	0.229	0.042	0.173	0.461
	6	0.078	-0.037	0.022	0.069	0.369
	7	0.073	-0.061	0.066	0.181	0.135
	8	0.100	0.129	-0.108	0.099	0.231
	9	-0.066	-0.121	-0.142	-0.076	-0.002
	Large	-0.133	-0.254	-0.173	-0.325	0.055
p-value		0.343	0.371	0.784	0.171	0.197

.

Table IX

Pricing Errors and p-values

The table shows the pricing errors and p-values for the SSD-TFS, and TSD-TFS tests, using the CRSP all-share index relative to the ten long-term reversal portfolios in the full sample and four non-overlapping subsamples.

Panel A: SSD-TFS						
		Long-term reversal portfolios				
		Jan 1931-Dec 2002	Jan 1931-Dec 2002	Jan 1931-Dec 2002	Jan 1931-Dec 2002	Jan 1931-Dec 2002
Pricing errors	Small	0.203	0.748	-0.109	0.374	0.347
	2	0.093	0.172	0.107	0.132	0.580
	3	0.152	0.280	0.134	0.263	0.467
	4	0.066	0.017	0.088	0.230	0.265
	5	0.105	0.166	0.079	0.128	0.349
	6	0.074	-0.053	0.028	0.104	0.351
	7	0.072	-0.101	0.025	0.140	0.289
	8	0.053	0.133	-0.117	0.081	0.186
	9	-0.059	-0.115	-0.107	-0.140	0.127
	Large	-0.129	-0.205	-0.116	-0.398	0.055
p-value		0.148	0.105	0.694	0.065	0.638
Panel B: TSD-TFS						
		Long-term reversal portfolios				
		Jan 1931-Dec 2002	Jan 1931-Dec 2002	Jan 1931-Dec 2002	Jan 1931-Dec 2002	Jan 1931-Dec 2002
Pricing errors	Small	0.314	1.069	-0.119	0.416	0.349
	2	0.192	0.576	0.090	0.175	0.556
	3	0.231	0.529	0.133	0.279	0.432
	4	0.095	0.167	0.079	0.237	0.243
	5	0.146	0.349	0.063	0.152	0.335
	6	0.078	0.014	0.007	0.100	0.364
	7	0.083	0.016	0.016	0.157	0.268
	8	0.069	0.190	-0.104	0.075	0.177
	9	-0.071	-0.136	-0.104	-0.165	0.105
	Large	-0.162	-0.346	-0.102	-0.431	0.024
p-value		0.117	0.091	0.812	0.070	0.595

4.7 Concluding remarks

This chapter offers an overview of the efficiency of the market of various datasets as well as their support for TFS. At first sight, the results may seem not overly remarkable, as MV efficiency is accepted for a number of datasets. The only dataset which offers the maximum amount of support possible (reject MV efficiency, no rejection of SD efficiency, partial support for TFS) is the set of beta sorted portfolios.

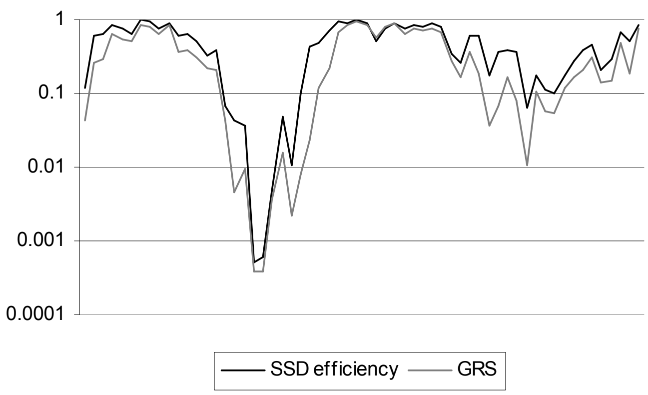
Yet the point remains that MV is incompatible with the regularity conditions as given by utility theory. Also, the MV specification is just one of many, with the test for TFS we can look at the robustness for specification errors in a much more comprehensive way. Partly, the results also reflect the period that was chosen. We tried to maximize the available data, which often resulted in the inclusion of the late twenties, which in many cases makes it harder to reject any theory. We have to accept the results the data brings us. However, the results form a substantial first step in gaining more insights by applying the newly developed SD-tests in empirical settings. In that respect, 3 issues warrant some further discussion:

Rolling window analysis

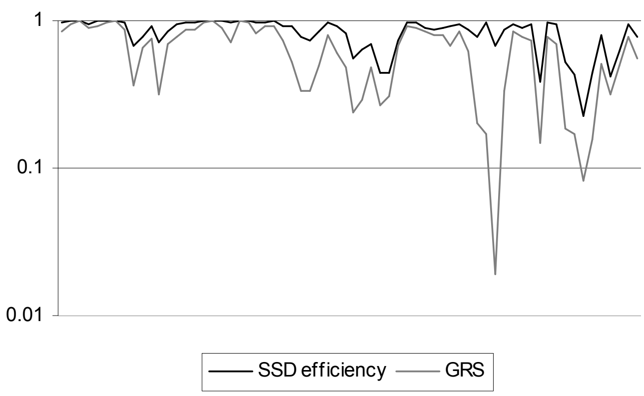
The fact that SD can detect efficiency in the beta sorted portfolios where the MV framework can't is not an accidental result; if we are to use rolling window analysis (use the SSD test for a period of 10 years, starting one year further each time) we see that this result is obtained in many subperiods in several datasets, but often drowns in the complete sample due to periods that are indeed compatible with variance as a risk measure. The results are in figure 5.

Also, we see that for the book-to-market and momentum portfolios the SSD p-value is in some periods actually below that of the GRS test, and by a substantial margin. This indicates that in those cases, the GRS test indeed chooses kernels which do not conform to the regularity conditions. A more detailed look into the subperiods seems desirable.

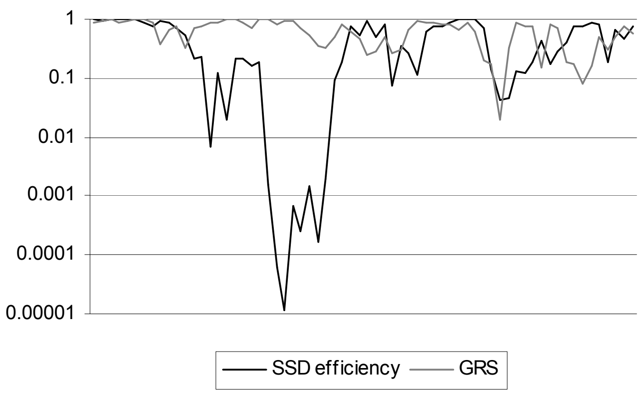
Panel A: p-values beta sorted portfolios



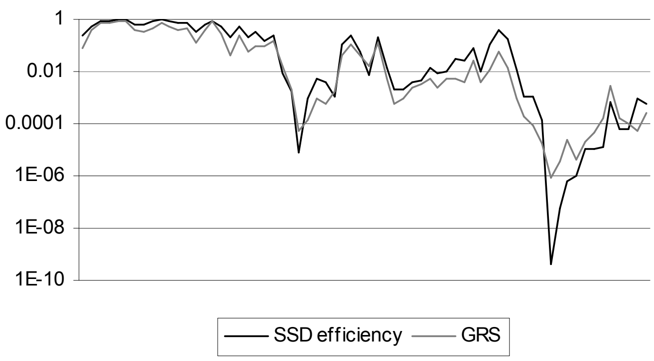
Panel B: p-values size sorted portfolios



Panel C: p-values value sorted portfolios



Panel D: p-values momentum portfolios



Panel E: p-values long-term reversal portfolios

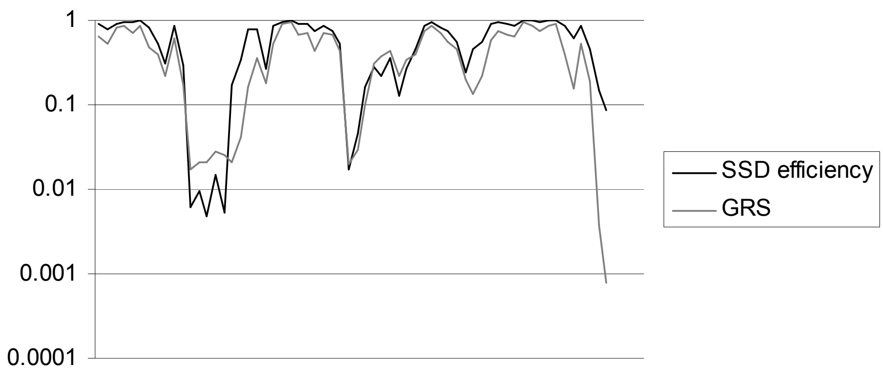


Figure 5

P-values rolling window analysis

The figure shows the p-values for the SSD over a period of 10 years, starting at the beginning of each sample and then advancing one year. The scale is logarithmic; panel A shows the beta sorted portfolios, panel B the size sorted portfolios, panel C the book-to-market sorted portfolios, panel D the momentum sorted portfolios and finally panel E the long-term reversal sorted portfolios. Note that the scales differ between the panels.

The 1946-1964 period

For most of the datasets analysed (as well as the non-reported short-term reversal set) the period of 1946 to 1964 is particularly hard to explain; even when there is a rejection of efficiency in other (sub)periods as well, the test statistic is high in this particular timeframe. Given the sensitivity of several datasets to the sample period (see above) this result is worrying. As it stands, we don't have an explanation for this which we can support with literature. Yet as the effect is present in the value (book-to-market) sorted portfolios, momentum portfolios and the beta and short-term reversal portfolios, it seems unlikely to be a coincidence. The table below gives details for the SSD efficiency test, but the results carry over to the TSD and GRS test as well, except for the book-to-market portfolios where GRS chooses a kernel which is not admissible under SD for this period.

Dataset	p-value 1946-1964	Test-statistic 1946-1964	Average test statistic[12]
Beta sorted	$1.8\cdot10^{-4}$	34.10	7.82
Momentum	$1.7\cdot10^{-4}$	34.17	25.56
ST reversal	$4.1\cdot10^{-5}$	26.06	21.83
Book-to-market	$3.7\cdot10^{-3}$	37.78	4.97

Inherent problems

It should be noted that certain problems are still inherent in this type of empirical research. By necessity I have to make a selection; below are the ones that we feel to merit further attention.

Firstly, one has to make a choice regarding the frequency of the data which is going to be used, which implicitly fixes an investment horizon, as it limits the moments in which trade is possible. We have chosen for a monthly frequency due to the fact that SD requires large amounts of data to give reliable results, yet at the end of the day there is no perfect solution to this problem.

Secondly, the model assumes that the utility specification remains constant over the entire sample period. This may be a reasonable assumption for a 5 or 10 year period, but when applied to 70 years of data (meaning that we almost surely have completely

[12] This is the average of the test statistic in 3 other subperiods of equal length, and is given for comparison.

refreshed the pool of actual investors, and probably many times over) it is an assumption one is less comfortable with. Is the risk attitude of the representative agent in the late twenties indeed the same as in the early fifties or the late seventies, or the turn of the millennium? Of course the representative agent itself is fictious, but the investors whose behavior and preferences we hope to aggregate into it aren't. Further research into methods using a more suitable way of dealing with this problem would be welcome. A first start would be to find suitable additional restrictions, as the flexibility of the kernel increases dramatically if it can change over time, thereby increasing the odds of an unwarranted non-rejection of the null-hypothesis.

Thirdly, the influence of foreign investors and exchange rates is ignored. At first sight one could object that foreign investors could be incorporated into the representative agent. It itself that is not the problem; the issue is that foreign agents have a different set of returns, as they are modified by exchange rates. If the exchange rate is just random 'noise' (statistically speaking) it would presumably only result in more uncertainty at the statistical analysis. However, for some markets the set of returns viewed through the 'lens of the exchange rate' may differ substantially from those observed by the native investors, and may be correlated with the exchange rate as well. A large negative return (i.e. a crash) may coincide with foreign investors liquidizing their assets and thereby sending the exchange rate into a similar predicament[13], which compounds the negative return. And SD analysis often tends to point to tail risk, meaning that it are the very large negative returns which particularly influence investors.

[13] Of course, vise versa is possible too. These effects may differ among markets; for instance the US market (which we analysed in this chapter) has a nominally large foreign component, but also very large investments from the US itself. Yet for Asian countries, there is at least a substantial body of anecdotal evidence that the flight of foreign capital may have (had) a profound effect along the liens described.