Evaluating Portfolio Performance with Stochastic Discount Factors — **Source link** 🔗

Magnus Dahlquist, Paul Söderlind

Related papers:

- Portfolio Performance Measurement: Theory and Applications

- Measuring Fund Strategy and Performance in Changing Economic Conditions

- Performance Evaluation with Stochastic Discount Factors

- Large sample properties of generalized method of moments estimators

- Implications of Security Market Data for Models of Dynamic Economies

Evaluating Portfolio Performance with Stochastic Discount Factors

Magnus Dahlquist
Duke University,
Stockholm School of Economics and CEPR

Paul Söderlind*
Stockholm School of Economics and CEPR

August 31, 1998

Abstract

This paper provides evidence on the use of stochastic discount factors in the evaluation of portfolio performance. First we discuss evaluation in this setting, and relate it to traditional mean-variance analysis. We then use Monte Carlo experiments to examine the small sample properties of various generalized method of moment estimators. Overall, the test statistics are fairly well described by their asymptotic distributions, even though serious size distortions are found in some cases. The simulations also show that a significant excess return, or a long sample, is needed to reject neutral performance. Finally, we offer an evaluation of Swedish-based mutual funds. The conditional

evaluation indicates that funds may have had non-neutral performance over the sample period as revealed by the predictability of the unconditional performance measure.

1

1 Introduction

Performance evaluation has recently received increased attention. Apart from the obvious interest for investors to evaluate portfolio strategies, this can be traced to at least two distinct developments in the asset pricing literature.

The first development is the use of efficient benchmark portfolios. Performance evaluation is essentially a question of comparing the risk-adjusted return on one portfolio with certain benchmarks. It is, however, well known that the use of inefficient benchmarks can cause ambiguity in an evaluation. The use of two inefficient benchmark portfolios could, for instance, reverse the ranking of passive portfolios, and portfolios which have superior performance using one benchmark may have inferior performance with another benchmark.[1] The second development is the use of conditional information variables in tests of asset pricing theories. Traditional approaches to performance measurement are unconditional, in the sense that average returns are being used to estimate the expected performance. These measures do not account for time-variation in expected returns and risk, which can mask truly inferior or superior performance.[2]

(Bansal and Harvey 1996), (Chen and Knez 1996), and (Farnsworth, Ferson, Jackson, and Todd 1996) combine these developments, and implement the performance measures in a conditional setting using efficient benchmarks. The evaluations are done in the context of (Hansen and Richard 1987), (Hansen and Jagannathan 1991), and (Snow 1991). Basically a stochastic discount factor (SDF) which prices all benchmarks (and is in this sense efficient) is constructed and then used in evaluations. The methodology is related to the recent spanning tests in (De Santis 1995), and (Bekaert and Urias 1996).

In this paper we provide further evidence on the use of stochastic discount factors in an evaluation of portfolio performance. Our paper is close in spirit to the work by (Chen and Knez 1996) which deal with the theoretical strengths and weaknesses in SDF performance measures. We extend the interpretation of the performance measures, and relate them to the traditional mean-variance analysis. We focus on the implementational aspects of the approach. Using simulations we examine the properties of the measures

---

[1] For further elaborations on this, see, for instance, (Roll 1978), and (Dybvig and Ross 1985).

[2] (Ferson and Schadt 1996) develops this reasoning, and extends the traditional measures of performance and market timing by using predetermined information variables.

in small samples. We consider SDFs with and without positivity imposed. We allow the benchmarks to be either fixed-weight strategies or dynamic strategies formed via available public information. Moreover, we consider unconditional as well as conditional evaluations.

In the case without positivity imposed on the SDF, there is an appealing duality between the performance measure and the traditional mean-variance analysis. In fact, the test of neutral performance can be interpreted as an intersection test. That is, the evaluation boils down to a test of whether additional assets (here mutual funds) expand the mean-variance frontier. The intersection test becomes equivalent to an evaluation of a fund's pricing error with respect to an ex post efficient mean-variance portfolio (an 'alpha'). Following (Chen and Knez 1996), we also consider evaluation measures where positivity on the SDFs is imposed. The positivity of the SDF prevent situations where one can find arbitrage opportunities, but there is no enlargement of the mean-variance frontier. Moreover, the SDFs are efficient by construction (as in (Hansen and Jagannathan 1991)), and require only the returns on the benchmark assets. The analysis does not rely on a specific model, and hence not on its accuracy.

Another strength of the approach is that estimation and testing can be done with the generalized method of moment (GMM) of (Hansen 1982). This means that specific distributional assumptions of the asset returns are not required, and we do not need to work in a normal iid setting. We can handle both conditional heteroskedasticity and serial correlation in pricing errors. Moreover, it is straightforward to incorporate conditional information, and the non-linearities (via positivity constraints on the SDFs) can be dealt with.

This framework is used in a comparison of actively managed mutual funds with some benchmark portfolios. Do fund managers possess information that provides superior returns or can simple portfolios price the returns of mutual funds? Does the incorporation of conditional (public) information via scaled returns matter? Given the interest of these questions, can our measures really distinguish superior performance from neutral performance? How do the empirical small sample distributions for the evaluation statistics look like under the null and alternative hypotheses? These are some of the questions we address.

We discuss the use and implementation of stochastic discount factors in a portfolio evaluation in Section 2. We also relate the evaluation to the traditional mean-variance analysis. We especially highlight the importance of getting the mean of the SDF right. Otherwise, the evaluation could be un-

dertaken at an arbitrary (and misleading) point in the mean-variance space. We also show how to select predetermined information variables in order to undertake a correct conditional evaluation. In Section 3 we describe data on Swedish benchmarks, instruments, and funds which we employ in the study. Monte Carlo experiments are used in Section 4 to examine the small sample properties of the GMM estimators. Both size and power properties are characterized using various GMM approaches. Overall, the test statistics are well described by their asymptotic distributions, even though serious size distortions are found in conditional evaluations. The simulations also show that a significant excess return, or a long sample, is needed to reject neutral performance. In Section 5 we apply the method to Swedish data. We undertake an unconditional as well as a conditional evaluation. The conditional evaluation indicates that funds may have had non-neutral performance over the sample period, as revealed by the predictability of the unconditional performance measure. Finally, we conclude in Section 6.

2   Stochastic Discount Factors and Performance Evaluation

We use the same basic approach to testing fund performance as (Chen and Knez 1996). We first estimate a stochastic discount factor (SDF) which 'prices' some benchmark portfolios, and then use GMM to test if it also prices the funds. The SDF is estimated from asset returns as in (Hansen and Jagannathan 1991), and we compare the results from SDFs with and without positivity imposed.

This section shows how to set up the GMM moment conditions to guarantee that simple trading rules are assigned neutral performance, both unconditionally and conditionally. The simple trading rules we consider are either fixed-weight portfolios or dynamic portfolios where the portfolio weights are allowed to depend linearly on some publicly available information.

Without positivity of the SDF, the GMM-SDF test is equivalent to investigating if the mean-variance frontier is expanded by adding funds. In this case, the main reason for using the GMM-SDF test is that it is easy to handle non-normal and non-iid pricing errors. Furthermore, to our knowledge, there is no alternative to the GMM-SDF test if we want to impose positivity of the SDF, which may have some advantages as discussed by (Chen and

4

Knez 1996). For instance, a fund may offer arbitrage opportunities without expanding the mean-variance frontier, and this will (sampling errors apart) be picked up by a positive SDF. Similarly, a positive SDF also guarantees that a fund which does systematically better than another fund is assigned a higher performance.

2.1 Performance Evaluation and Benchmark Portfolios

Investors form portfolios of certain 'benchmark' assets. The one-period gross return on such a portfolio is

$$R_t = w'_{t-1} R_{1t}, \text{ with } w'_{t-1} \mathbf{1} = 1, \tag{1}$$

where R_{1t} is a vector of the one-period gross returns on benchmark assets, and w_{t-1} the portfolio weights chosen in $t-1$. We assume that the information set and preferences lead investors to use simple trading rules where w_t could be any linear function of some publicly available information, x_t,

$$w_t = \delta' x_t, \text{ such that } w'_t \mathbf{1} = 1. \tag{2}$$

We will focus on two cases. First, with *fixed-weight benchmark portfolios*, x_t is just a constant. Second, with *dynamic benchmark portfolios*, x_t is a vector with a constant and augmented with some information variables which have been proven to predict stock market data. There must be differences in the information set and preferences in order to motivate the differences in trading rules. In spite of this, we assume that the time series process of the benchmark asset returns, R_{1t}, is exogenous and the same for both cases.

Let z_{t-1} be a vector of instruments known at $t-1$. The law of one price implies the existence of an SDF, m_t, such that

$$\mathrm{E}\lambda_{1t} = \mathbf{0}, \text{ with} \tag{3}$$

$$\lambda_{1t} = z_{t-1} R_t m_t - z_{t-1}. \tag{4}$$

In some cases we also impose the restriction that the SDF must be positive, as suggested by (Hansen and Jagannathan 1991).[3]

---

[3]In a conditional asset pricing model, the absence of no arbitrage opportunities (non-negative payoffs with negative prices) implies that there must be some SDF which is positive almost surely.

Suppose we now allow investors to take an infinitesimal position in a fund. The fund invests in the same benchmark assets as investors, but with potentially more sophisticated trading rules, for instance, because of superior information. We study if these trading rules add anything to the opportunity set of investors. The flavor of the testing approach is to first find a SDF which 'prices' the benchmark strategies, that is, satisfies (3)-(4). Then, we see whether it also prices the fund. That is, if the following conditions are satisfied

$$\mathrm{E}\lambda_{2t} = \mathbf{0}, \text{ with} \tag{5}$$

$$\lambda_{2t} = z_{t-1}R_{2t}m_t - z_{t-1}. \tag{6}$$

Different choices of the instruments give tests of different aspects of the fund performance. We focus on two cases. First, we get an *unconditional performance measure* by letting z_t be a constant. It is straightforward to show that $\mathrm{E}\lambda_{2t}$ is then the average return of the fund minus what the SDF implies for the risk free rate and the risk premium.[4]

Second, we get a *conditional performance measure* by using both a constant (nests the first case) and some information variables in z_t.[5] The element in $\mathrm{E}\lambda_{2t}$, corresponding to the constant, is still a measure of average performance, while the other elements capture the predictability of the fund return. For simplicity, suppose the average performance is zero. It is then straightforward to show that the remaining elements in $\mathrm{E}\lambda_{2t}$ are the covariances between the predetermined variables and the period t performance of the fund.[6]

---

[4]If z_{t-1} is a constant, then (3) can be written $\mathrm{E}R_t - 1/\mathrm{E}m_t = -\mathrm{Cov}(R_t, m_t)/\mathrm{E}m_t + \mathrm{E}\lambda_{1t}/\mathrm{E}m_t$. Since m_t prices R_t, $\mathrm{E}\lambda_{1t} = 0$, so any risk-free rate has the mean $1/\mathrm{E}m_t$ and the risk premium is captured by the covariance term. (5) can be written on the same form, which gives the result in the text.

[5](Ferson and Schadt 1996) conduct a 'conditional' evaluation in a linear beta framework, by allowing the betas to vary with predetermined information variables. This corresponds to our unconditional evaluation of dynamic strategies.

[6]Let λ_{1t}^{uc} and λ_{1t}^c be the elements in λ_{1t} corresponding to the constant (unconditional evaluation) and an information variable z_{t-1}^c (conditional evaluation), respectively. Clearly, $\lambda_{1t}^{uc} = R_t m_t - 1$ and $\lambda_{1t}^c = \lambda_{1t}^{uc} z_{t-1}^c$, so $\mathrm{E}\lambda_{1t}^c = \mathrm{Cov}(\lambda_{1t}^{uc}, z_{t-1}^c) + \mathrm{E}\lambda_{1t}^{uc}\mathrm{E}z_{t-1}^c$. Since m_t prices R_t, both $\mathrm{E}\lambda_{1t}^{uc}$ and $\mathrm{E}\lambda_{1t}^c$ are zero, so average performance of R_t is zero and the performance in t, λ_{1t}^{uc}, is not correlated with z_{t-1}^c. (5) can be written on the same form, which gives the result in the text.

2.2 Construction of the SDFs and GMM Testing

We need to estimate SDFs which price the returns from the two different classes of simple trading rules discussed above. This can be done by replacing (4) by

$$\lambda_{1t} = (z_{t-1} \otimes x_{t-1} \otimes R_{1t}) m_t - z_{t-1} \otimes x_{t-1} \otimes \mathbf{1}, \tag{7}$$

and then finding an m_t such that $\mathrm{E}\lambda_{1t} = \mathbf{0}$. As shown in Appendix A, *any* portfolio with weights which are linear in the information variables x_{t-1}, and summing to one in *every* period, will then be assigned a neutral performance.

We will later show that it is important to tie the mean of the SDF to a reasonable value. This can be achieved by adding a moment condition related to an asset whose return, R_{ft}, is a suitable level for a hypothetical risk free rate

$$\mathrm{E}\lambda_{ft} = 0, \text{ with} \tag{8}$$
$$\lambda_{ft} = m_t - 1/R_{ft}. \tag{9}$$

We follow (Hansen and Jagannathan 1991) and (Chen and Knez 1996) and construct the SDFs from the sample of $\{1, R_{1t}, x_{t-1}, z_{t-1}\}$. Without positivity, m_t is a linear function of the cross product of the four vectors. Positivity is imposed by cutting off the linear function at zero in an option-like fashion. In principle, the testing approach could be to choose the vector γ in

$$m_t^* = \left[\begin{array}{cc} 1 & (z_{t-1} \otimes x_{t-1} \otimes R_{1t})' \end{array} \right] \gamma, \text{ or} \tag{10}$$
$$m_t^+ = \max \left(\left[\begin{array}{cc} 1 & (z_{t-1} \otimes x_{t-1} \otimes R_{1t})' \end{array} \right] \gamma, 0 \right) \tag{11}$$

to make the sample average of λ_{1t} in (7) and λ_{ft} in (9) equal to a vector of zeros.[7] (This is possible since there are as many elements in γ as there are equations.) The series of estimated m_t^* or m_t^+ could then be used to test if (5) holds. Noteworthy is that if there are common variables in the z_{t-1} and x_{t-1} series, *only* the unique cross-products in $z_{t-1} \otimes x_{t-1}$ have to be used in (10) and (11) to price the benchmarks under the null hypothesis.

However, to take into account the sampling error in the estimated SDF, we employ the following one-pass procedure instead. The vector γ is estimated

---

[7](Bansal and Viswanathan 1993) consider SDFs which are non-linear in asset returns, and (Glosten and Jagannathan 1994) discuss non-linearities for performance evaluation.

by minimizing a quadratic loss function of the sample averages of not only λ_{1t} and λ_{ft}, but also λ_{2t}

$$\left(\frac{1}{T} \sum_{t=1}^{T} \begin{bmatrix} \lambda_{1t} \\ \lambda_{ft} \\ \lambda_{2t} \end{bmatrix} \right)' W \left(\frac{1}{T} \sum_{t=1}^{T} \begin{bmatrix} \lambda_{1t} \\ \lambda_{ft} \\ \lambda_{2t} \end{bmatrix} \right). \tag{12}$$

The weighting matrix W is positive definite, but need not be the inverse of the covariance matrix of the moment conditions as is most often the case. Each element in $\sum_{t=1}^{T} \lambda_{2t}/T$ is an overidentifying restriction, and we can test if they are all satisfied by using the χ^2 test in (Hansen 1982) (see Appendix B for details). Without positivity, we have a linear model which can be solved analytically, but positivity introduces a non-linearity which requires a numerical optimization algorithm.

2.3   The Relation to Mean-Variance Analysis

(Hansen and Jagannathan 1991) show the relation between the mean-variance frontiers for asset returns and for the SDF *without positivity*. We use this framework to discuss how the GMM-SDF test without positivity relates to traditional mean-variance tests. We focus on the unconditional case with fixed-weight benchmark portfolios, $x_t = z_t = 1$, but the results can also be applied to the managed portfolio returns, $z_{t-1} \otimes x_{t-1} \otimes R_{1t}$ and $z_{t-1} \otimes R_{2t}$.

(Huberman and Kandel 1987) construct a test of whether the mean-variance frontier of R_{1t} intersects the mean-variance frontier of R_{1t} and R_{2t}. Suppose the fund return is

$$R_{2t} = a + BR_{1t} + \varepsilon_t, \text{ with } \mathrm{E}\varepsilon_t = 0 \text{ and } \mathrm{E}R_{1t}\varepsilon_t' = \mathbf{0}. \tag{13}$$

(Huberman and Kandel 1987) show that intersection is equivalent to the restriction

$$a = R^0(\mathbf{1} - B\mathbf{1}), \tag{14}$$

where R^0 is the zero beta rate associated with the portfolio at the intersection point. The relation to the GMM-SDF test of $\mathrm{E}\lambda_{2t} = 0$ is seen by using (13) to substitute for R_{2t} in (6). If the fund has neutral performance, then

$$a = \frac{\mathbf{1} - \mathrm{Cov}(m_t^*, BR_{1t})}{\mathrm{E}m_t^*} - B\mathrm{E}R_{1t}. \tag{15}$$

8

It is straightforward to show that this is equivalent to (14) (see, for instance, (Ferson 1995) and (Bekaert and Urias 1996)). The GMM-SDF test is thus an intersection test.

It is then intuitively clear that the GMM-SDF measure of performance, $\mathrm{E}\lambda_{2t}$, must be a kind of Jensen's alpha. To show that, combine the following facts: *(i)* (5) implies a beta relation between $\mathrm{E}R_{2t}$ and the SDF; *(ii)* the SDF portfolio return, $R_{bt}^* = m_t^*/\mathrm{E}m_t^{*2}$, is on the mean-variance frontier with a risk free asset ((Chamberlain and Rothschild 1983)); and *(iii)* the tangency portfolio, R_{Tt}^*, is perfectly correlated with the SDF portfolio. Together, these facts imply

$$\frac{\mathrm{E}\lambda_{2t}}{\mathrm{E}m_t^*} = \mathrm{E}R_{2t} - \frac{1}{\mathrm{E}m_t^*} - \frac{\mathrm{Cov}(R_{2t}, R_{Tt}^*)}{\mathrm{Var}(R_{Tt}^*)}\left(\mathrm{E}R_{Tt}^* - \frac{1}{\mathrm{E}m_t^*}\right), \qquad (16)$$

which is a Jensen's alpha computed against an *efficient* tangency portfolio. Hence, the fact that m_t^* prices the benchmarks correctly, yields an evaluation versus an efficient benchmark asset. The pitfall of evaluating against an inefficient asset is thus avoided. Finally, recall that this duality result holds only for the SDF without positivity.

We illustrate this in Figures 1.a-b, by plotting the mean-standard deviation frontiers of our data (to be described later). In Figure 1.a, the frontier for the benchmarks, R_{1t}, is the thick solid hyperbola. Adding one hypothetical fund with neutral performance gives the thick dashed hyperbola. These two frontiers intersect at the tangency portfolio, R_T^*, (marked by a circle), corresponding a risk free return, $1/\mathrm{E}m_t^*$, of about 10%. In Figure 1.b the fund has been given a 5% excess return, so the two frontiers no longer intersect at the tangency portfolio.

The moment condition (8) was included to tie the mean of the SDF to a reasonable value. Figure 1.b illustrates why this is necessary. Suppose we would incorrectly estimate $1/\mathrm{E}m_t^*$ to 28% (marked by R^0 on the vertical axis). This would imply a tangency portfolio at R_T^0 (marked by a triangle). Evaluated at this point, the fund with 5% excess return appears to have neutral performance. (Chen and Knez 1996) use only stock market data to estimate the SDF. In our data set, this leads to an implied risk free rate which differs significantly between different sets of benchmark assets, and the estimated rate is often unreasonable. This is the reason why we include the moment condition (8) and use the return on a short-term bond portfolio to capture the level of the risk free rate.[8]

---

[8](Farnsworth, Ferson, Jackson, and Todd 1996) also discuss the importance of getting

3 Data

In this section we describe the benchmarks, information variables, and funds we use in the Monte Carlo simulations and the performance evaluation of the funds. We employ a weekly Swedish data set covering the period from January 1986 to December 1995.

Ideally in an evaluation, we would like to have access to all the assets which comprise the different managers investment opportunity set. However, practical problems (like the handling of large econometric systems) prevent us from that. For this reason we consider various portfolios of assets as benchmarks, and evaluate the performance of the funds relative to these portfolios.

The first type of assets we include are truly passive buy-and-hold stock portfolios. The raw data used for the construction of the portfolios includes weekly observations of Swedish stocks listed on the official A-list at the Stockholm Stock Exchange (SSE), and are obtained from *Findata*. Stocks on the A-list correspond to on average 97% of the total market value and 96% of the total trading volume at the SSE over the sample period.[9] We sort the firms into five industry groups: manufacturing, wood and pulp, retail services and constructing, bank and insurance, and miscellaneous. We also consider a general portfolio based on all firms. The stocks are selected based on liquidity (the value of trade) and size (market capitalization value). The portfolio weights display initial market capitalization, and are rebalanced only once a year (to reflect changes in the market capitalizations of the individual stocks). We also employ a data set of five portfolios formed on the basis of market capitalization (size). A detailed description of the construction of the portfolios is given in (Dahlquist and Söderlind 1997).

The second type of asset we consider is a proxy for cash. We use the return on a short-term bond portfolio as a proxy for cash. It is constructed from a series of yields on a six-month Treasury bill.

To allow for dynamic trading strategies and to study conditional performance, we use a set of predetermined economic variables, which are meant to capture the state of the economy. There is considerable empirical evidence that the term structure of interest rates contain information about future

---

the mean of the SDF right.

[9]Stocks are traded under different listings, and the official A-list has the most stringent listing requirements on the records and stability of the company, as well as on the distribution and the liquidity on the company's stock.

economic variables (including stock returns). Two variables are used to capture the shape of the yield curve. The first is a 'level' variable based on the annualized yield of the six-month T-bill. It is stochastically detrended by subtracting a lagged moving average (two months) in order to decrease the persistence in the original yield. The second variable is the return on a bond portfolio with a maturity between 3 to 5 years in excess of the return on the bill portfolio. This is basically capturing changes related to the 'slope' of the yield curve, and will be referred to as the bond excess return. Moreover, the two variables are scaled so that their expected values are equal to one, and with variances about the same magnitudes as the original portfolio returns. The data series are obtained from *Datastream*. The lagged return on the general stock portfolio is used as a third information variable.

Table 1 reports the distributional characteristics of the industry portfolios, the bond portfolio, as well as the instrument variables (not scaled). The average net returns of the stock portfolios are in the range from 14% to 21% per year. The standard deviations of the stock portfolios are between 18% and 30% per year. The low average return of the short-term bond portfolio (about 10% per year) is associated with a low volatility of 1.3%.

Furthermore, the table shows the coefficients of skewness and excess kurtosis of the portfolio returns. If they are normally distributed, then they should be equal to zero. There is no evidence of skewness – some are positive and some are negative. However, there is strong evidence of excess kurtosis indicating that the peakness is higher and the tails are thicker than a normal distribution. This is economically and statistically significant as revealed by the magnitudes of the coefficients and the p-values (not reported).[10] A joint hypothesis of normality (not reported) can clearly be rejected at the 5% level for the portfolios.

The first-order serial correlations of the weekly returns are also reported in Table 1. They all have correlation coefficients greater than 0.11. We computed standard errors as in (Richardson and Smith 1994), allowing for persistence in the variances, which reveal that they all have p-values about 5%.

Finally, the estimated moments for the two term structure variables are reported. Notable is the high persistence in the level variable, and the contrasts in the two variables – indicating that they capture various aspects of

---

[10]The moments in the table are estimated within a GMM system, which gives us standard errors (and p-values) and different hypotheses can be tested.

the yield curve. The general (buy-and-hold) stock portfolio shows similar characteristics as the industry portfolios.

We have obtained data on mutual funds from *Findata*. They are available as net asset values per share adjusted for management fees, and are collected on the same weekday as the benchmarks. Dividends are reinvested, by purchases of shares, in the fund. The mutual funds have some load and sale charges which are not included. We choose to study funds which invest mainly in Swedish stocks. The benchmarks above should therefore correspond to the investment styles of the funds. The data set includes only funds which existed at the end of the sample period, and potential survivorship biases are not dealt with. The funds have different inclusion dates, and we mainly present results for the period July 1991 to December 1995 (235 observations). This gives us 24 funds in total. However, we also report the results for a sub-sample of 6 funds, covering January 1987 to December 1995 (465 observations)

The funds show similar characteristics as the benchmarks (not reported). The average returns and standard deviations are in the same ranges as the stock portfolios. There is little evidence on skewness different from zero, but significant excess kurtosis and serial correlation.

4   A Monte Carlo Experiment

4.1   Motivation

GMM is a very flexible approach which allows us to handle both non-iid and non-normal pricing errors and positivity restrictions. In contrast, traditional tests for mean-variance intersection/spanning (see, for instance, (Huberman and Kandel 1987)) rely on iid normally distributed errors and linearity.

The asymptotic properties of GMM estimators and test statistics are well known, but less is known about the finite sample properties of GMM-based performance tests. (Bekaert and Urias 1996) provide some simulation evidence for tests of unconditional mean-variance spanning. They find that it may be important to account for serial correlation and that the properties of large systems are very unsatisfactory. However, their simulations leave several important questions unanswered. First, how does a prespecified weighting matrix (motivated below) compare with the usual 'efficient' weighting matrix? Second, do the tests have the power to detect superior

performance? Third, can we rely on overlapping data? These are some of the questions we try to answer by performing a series of Monte Carlo simulations. Moreover, we extend their results to a conditional setting where data is consistently generated to allow for time-variation in expected returns via information variables. We also use the opportunity to compare the small sample properties of intersection and spanning tests.

4.2   Setup

We estimate a time series process from data on the information variables and the benchmarks described in Section 3. The estimated process is then used in the simulations.

We want a parsimonious modelling, but still capture the time-variation in the first and second moments. The information variables are therefore modeled as a VAR(1) system with conditionally heteroskedastic errors following a BEKK GARCH process (see (Engle and Kroner 1995)). The benchmark returns are generated as linear functions of the information variables, and the fund returns in accordance with Huberman and Kandel's (1987) restriction (13) under the assumption of mean-variance spanning, $B\mathbf{1} = \mathbf{1}$. The reason for generating data under spanning is that it certainly implies intersection (which we test for), and that we want to compare intersection and spanning tests. A constant is added to the fund return when we simulate a fund with excess return. More details are found in Appendix C.

For each experiment, we construct 3,000 Monte Carlo samples with 250 or 500 observations. A sample size of 250 (500) corresponds to approximately 5 (10) years of weekly data. From the simulations, we compute the empirical (small sample) distribution of the test statistic of neutral performance, that is, of the overidentifying restrictions that $\mathrm{E}\lambda_{2t} = \mathbf{0}$.[11]

The number of simulations quickly gets very large. We therefore show results from just a representative subset of the simulations, and make a few remarks about the others.

---

[11]We have also looked at the sample averages of the performance mesure itself. We find no evidence of biases and the distribution appears symmetric and similar to a normal distribution.

4.3 Choice of Weighting Matrix

In this section we study how the small sample properties of the GMM-SDF test are affected by the choice of weighting matrix W in (12). We compare four versions of 'efficient' weighting matrices (giving the smallest asymptotic variance matrix) with each other and with a prespecified, 'inefficient,' weighting matrix.

The first efficient weighting matrix is a two-step estimator. Start by setting W to an identity matrix and estimate the model parameters, and use these to estimate the covariance matrix of the sample averages of the moment conditions. Then, use the inverse of the covariance matrix as the new weighting matrix and estimate the model parameters again. The second efficient weighting matrix continues from the two-step estimator by reestimating the covariance matrix and the parameters until the loss function value converges – an iterated estimator.

The covariance matrix of a sample average is a sum of the autocovariances of the series. Estimators of this covariance matrix differ with respect to how many autocovariances that are used and which coefficients they have in the sum. In the first two efficient weighting matrices we assume a first-order autocorrelation (since returns show no partial autocorrelations beyond the first lag), and the coefficients follow Bartlett's 'tent shaped' scheme as in (Newey and West 1987). The third and fourth efficient weighting matrices are similar to the second, except that they use the automatic selection of the order of autocorrelation suggested by (Newey and West 1994) and (Andrews 1991), respectively.

(Cochrane 1997) argues that a prespecified weighting matrix may give more robust results and that it facilitates comparison between estimations (here funds). We therefore try the inverse of the second moment matrix of (managed) returns suggested by (Hansen and Jagannathan 1997). This matrix may be of particular interest, since the minimized GMM loss function value can then be interpreted as the distance between the SDF estimated with the overidentifying restrictions and the SDF which prices all assets (including the fund).

We use three tests in the paper: unconditional performance with fixed-weight benchmarks, unconditional performance with dynamic benchmarks, and conditional performance with dynamic benchmarks. For the first two tests, the different weighting matrices give virtually the same result. In contrast, for the test of conditional performance with dynamic benchmarks,

the choice of weighting matrix matters.

The test of conditional performance with dynamic benchmarks is a much larger system of equations than the first two tests (the number of estimated parameters are 6, 21, and 51, respectively). Simulations reveal that this difference is important in small samples, but get less and less important as the sample size increases. The number of observations per parameter, the saturation ratio, seems to be a key feature of these, and several subsequent, simulations.[12] Here, it is 51, 20, and 9 for the three tests, respectively. A value below 10 is often considered as an indicator of potential problems.

Figure 2.a shows the empirical probability of rejecting a true null hypothesis of neutral performance (the 'size' of the test) as a function of the level of the critical value from the asymptotic distribution (nominal size). The closer the curves are to a 45 degree line, the closer are the small sample distributions to the asymptotic (chi-square) distribution.

The prespecified weighting matrix is slightly better than the 2-step estimators, which are all very similar. However, the iterated estimator (in this case based on Bartlett's scheme, but this is not important) has clearly a smaller size distortion than the other estimators. It can be shown that the differences between the weighting matrices are smaller when the sample is longer.

Suppose we forget about the $x\%$ critical values from the asymptotic distribution, and instead take those values that give rejections in $x\%$ of the simulations in Figure 2.a. We now apply these 'empirical critical values' on simulated data where the fund has 5% excess return per year, to see how often the false hypothesis of neutral performance is rejected (the 'power' of the test). Figure 2.b show – once we use the empirical critical values – that all estimators do equally well.

This simulation exercise suggests that it might be important to use small sample (simulated) critical values. However, once that is done, all the estimators considered here appear equally good. The prespecified weighting matrix has some practical advantages, so this is our main choice for the subsequent simulations and the evaluation of Swedish mutual funds. However, none of

---

[12]The saturation ratio (measured as the number of observations per parameter, including the unique ones in the weighting matrix) discussed in (Gallant and Tauchen 1989) can be seen as a guide to whether the dimension of the econometric system is too large. (Ferson and Foerster 1994) improve the accuracy of GMM standard errors by adjusting for the total number of parameters. The saturation ratios in our setting are shown in Appendix D.

the results depend on this choice.

4.4   Power Issues

In this section we investigate whether it is possible to reject the hypothesis of neutral performance when it is false.

In Figure 3 we compare the three tests used in the paper on a sample of 250 weekly observations. Figure 3.a show that there are small overrejections in the two unconditional tests. For instance, a 5% critical value from the asymptotic distribution, which is $\chi^2(1)$, generates rejections in 6% to 7% of the simulations.[13] The overrejection is more pronounced in the conditional test: the 5% asymptotic critical value, from a $\chi^2(4)$ table, gives rejections in 13% of the simulations.

Figure 3.b show that the power of the two unconditional tests is not too impressive, even if we allow the fund to have 5% excess return per year. The power of the conditional test is even lower. The basic reason is that the fund returns are very volatile, and the 'managed fund' returns, $z_{t-1} \otimes R_{2t}$, even more so. To illustrate this effect, disregard the risk-adjustment and suppose that a fund has 5% excess return in a sample of τ years. If the excess return is serially uncorrelated and has a standard deviation of 10% per year, then it takes more than 10 years of data to reject neutral performance using a 10% critical value (solving $5/(10/\sqrt{\tau}) = 1.64$ gives $\tau \approx 10.8$).

In Figure 4 we take a closer look at how the sample length and the excess return affects the size and the power. In this case, we show results for the unconditional test with dynamic benchmarks. Figure 4.a shows that the size is not affected very much by increasing the sample length from 250 to 500 weekly observations. This is true also in the unconditional test of fixed-weight benchmarks. In the conditional test, we notice that size distortions decrease substantially for the 2-step estimators and the prespecified weighting matrix, but not for the iterated estimator. The relative advantage of the iterated estimator for the conditional test on a small sample, shown in Figure 2.a, is therefore almost eliminated in the longer sample.

Figure 4.b shows that the power increases significantly when the excess return goes from 3% to 5% per year or when the sample length is increased

---

[13]The Monte Carlo results have sampling errors since they are based on a finite number of simulations (i.e. 3,000). A 10% Kolmogorov-Smirnov confidence interval of the distributions in the figures would be approximately ± 0.02 around the point estimate.

from 250 to 500. For instance, at the empirical 10% critical values, the power increases from 20% to 40% when the excess return goes from 3% to 5%. Similarly, increasing the sample length from 250 to 500 increases the power at 5% excess return from 40% to 70%. As a comparison, consider once again a t-test of an iid normal excess return. In this case, the power is expected to increase at roughly the same speed as the mean excess return increases from 3% to 5%, but at a slower speed as the sample length increases.[14] The other tests give similar results, except that the return to increasing the sample length is particularly large for the conditional test (where the saturation ratio is initially low).

This simulation exercise shows that it may be important to use simulated critical values, and that a considerable excess return or several years of data (or both) is necessary in order to reject neutral performance. This is especially true for the conditional test.

4.5 Overlapping Data

In this section we study the effect of the sampling frequency and the evaluation horizon. In general, statistical methods benefit from more information, but it could be the case that a high sampling frequency adds so much noise that it only makes things worse.

We have access to weekly data, so we can choose to evaluate either non-overlapping weekly returns or the more common monthly returns. If we use monthly returns, then there is a choice between sampling once a month (non-overlapping monthly returns) or once a week (overlapping monthly returns). GMM makes it straightforward to handle the serial correlation that is typically induced by overlapping data.

The results are illustrated in Figure 5. In this case, we show results for the unconditional test with dynamic benchmarks on a sample of 248 weeks, but the other tests give the same type of results.

Figure 5.a shows that non-overlapping weekly returns give fairly good size properties (this is the same curve as in Figures 2 and 4, which both used weekly returns). In contrast, monthly returns (sampled weekly or monthly) gives large size distortions, with rejections in 10%-12% of the simulations at

---

[14]If x is distributed as $\mathrm{N}(\mu, \sigma^2)$, the probability that the t-statistics of the sample mean exceeds 1.64 is $P = 1 - \Phi(1.64 - \sqrt{T}\mu/\sigma)$. Set $\{P, T, \mu\} = \{0.2, 250, 0.03\}$ and solve for σ. With this value of σ, P is 0.38 at $\{T, \mu\} = \{250, 0.05\}$ and 0.63 at $\{T, \mu\} = \{500, 0.05\}$.

the 5% asymptotic critical values. The power properties in Figure 5.b indicate that weekly data (either non-overlapping weekly returns or overlapping monthly returns) deliver somewhat higher power than monthly data.

This simulation exercise suggest that we should use either of the two weekly data series. It seems somewhat more straightforward to use non-overlapping data, so our main choice for the evaluation of the Swedish funds and the simulations (perhaps already noticed) is to use non-overlapping weekly returns.

4.6   Intersection versus Spanning Tests

In this section we round off the simulation experiments by a simple comparison of the intersection test we use and the spanning test used in (De Santis 1995) and (Bekaert and Urias 1996).

All portfolios on the mean-standard deviation frontier can be written as combinations of two portfolios on the frontier, so intersection at two distinct points is the same as spanning. This is the basis for the spanning test of (Bekaert and Urias 1996), which tests for intersection at two prespecified points (0% and 10%). We compare this with a test of intersection at one prespecified point (10%), which is a slight modification of our performance test (we no longer use (8) to tie down the intersection point.)

The results are illustrated in Figure 6. As before, we show results for unconditional performance with dynamic benchmarks on a sample of 250 weeks, but we get the same type of results in the other cases. The size is almost the same for the two tests, but the intersection test has somewhat better power. The intersection and spanning tests have both saturation ratios above 10 – enough to not show large size distortions. The power result could be related to the higher saturation ratio in the intersection test (it has only half as many moment conditions as the spanning test).

4.7   Positive SDF

The SDF without positivity gives a linear model with an analytical solution, while positivity introduces a non-linearity which requires a numerical optimization algorithm.

In the unconditional tests, it was quite straightforward to solve the non-linear minimization problem, and the small sample properties were very similar to those for the SDF without positivity reported above. We interpret

this as good news, since there are typically reasons to worry about the small sample properties of non-linear estimation methods.

However, in the conditional test, the solution algorithms (we tried quite a few) proved unreliable. In this case, there are 51 parameters to estimate from 55 moment conditions, which is a fairly large system of non-linear equations. The gradient based methods often got stuck at local minima, and other methods were too slow in order to perform a meaningful Monte Carlo experiment.

This simulation exercise shows that the small sample properties of the test with positivity of SDF are very similar to those of the SDF without positivity – at least in the case of an unconditional test.

4.8 Conclusions from the Simulation Exercises

The main results on the small sample properties of the GMM-SDF test are the following. First, asymptotic critical values should not be used, especially not when the number of observations per estimated parameter, the saturation ratio, is low. Second, it takes a considerable excess return and a long sample to reject neutral performance. Third, a prespecified weighting matrix does as well as the other weighting matrices – once empirical critical values are used. Fourth, non-overlapping weekly returns give better size and power than monthly returns. Fifth, the intersection test has better power than the spanning test. Sixth and finally, imposing positivity on the SDF is only feasible for the unconditional tests, but has no effect on the statistical properties.

We have only reported a subset of the simulations. In other simulations we have looked at the effect of using fewer/more benchmark assets, a shorter/longer sample, weaker/stronger instruments, combinations of the various experiments reported above, and different types of data generating processes. Given the previous results, these additional simulations did not reveal any news.

5 Evaluating Swedish Mutual Funds

We now evaluate the performance of a sample of Swedish mutual funds. The returns are non-overlapping weekly returns, and a prespecified weighting matrix (inverse of second moment matrix) is used – as suggested by the simulation results. We study mainly the sample period July 1991 to December

1995. This period is shorter than the total sample period, but includes more funds. We also report results for the period January 1987 to December 1995. First we evaluate the mutual funds using fixed-weight strategies. We then extend the analysis by incorporating dynamic strategies, and undertaking a conditional evaluation.

The stated goal of many funds is to provide investors with well diversified portfolios comprised of Swedish stocks. Our buy-and-hold stock portfolios are meant to describe the investment opportunity set, and should capture passive investment strategies of the funds (some properties of these portfolios are summarized in Table 1). The return on the short-term bond is used to tie down the mean of the stochastic discount factor, which was shown to be of importance.

We estimate the systems for all funds separately. For the unconditional evaluations, we initially estimate the models without imposing positivity on the SDF, and then we proceed to check whether the estimated SDF is always positive. If not, then we reestimate with positivity imposed. In the conditional evaluation with a large GMM system, however, we do not impose positivity because we ran into numerical problems.

What kind of results should we expect? Superior performance may be present if fund managers have information, which is not publicly available, and the skill to use it. The mutual fund returns we have (changes in net asset values) are adjusted for management fees which are not considered in the construction of the passive buy-and-hold portfolios, or in the dynamic strategies we allow for. Significant management costs would then affect the measures downwards. However, load charges and sales costs faced by the investors in the mutual funds are not incorporated in the net asset values. This would affect the measures upwards. In total, it is hard to see a systematic bias (in either direction) of these transaction/rebalancing costs. However, the data may suffer from survivorship biases. If it is the case that only the funds which have performed well during the studied period are in the sample, we can expect an upward bias in the performance measures.

We illustrate the results in figures. Consider first the unconditional evaluation, where the original benchmarks are assumed to reflect *fixed-weight strategies*. Figure 7.a shows the frequency of the statistic of the test for neutral performance (the overidentifying restriction). This is compared with the empirical density function under the null (which corresponds asymptotically

20

to the $\chi^2(1)$-distribution).[15] Empirical critical values from the simulations are also reported (the vertical dotted lines mark the 5% and 10% values). About 17% of the funds are to the right of the 5% critical value. This indicates that a somewhat larger portion than expected have non-neutral performance.

In Figure 7.b the distribution of the performance measure, $E\lambda_{2t}$, which utilizes a positive SDF, is reported. We see that most of the them are concentrated around zero but that some funds seem to have non-neutral performance. The distribution is somewhat skewed to the right, as indicated by the higher average (0.5%) compared to the median (0.1%) of the distribution. Out of 24 funds, 5 have absolute deviations (inferior or superior performance) which are greater than 5% (per year) from zero. Still, there is no general tendency in the distribution.

The benchmark portfolios can be seen as fixed-weight portfolios. We find that the average (risk-adjusted) performance is slightly positive. Can this be explained by dynamic asset allocation based on publicly available information? Alternatively, is the actual performance masked by variations in expected returns and risk-premia? Above we discussed *dynamic strategies* as a partial solution to this problem. We therefore consider dynamic strategies using the lagged level of the yield curve, lagged excess return on the bond portfolio, as well as the lagged return on the general stock portfolio as instruments. Put differently, we now analyze the unconditional performance against 5 managed portfolios. The results are shown in Figures 8.a-b. We can, however, see that the results are not very different. The distribution of the moment conditions for the funds in Figure 8.b is somewhat flatter with a mean and median of 0.4% and 0.1%, respectively. Again 5 funds have an absolute overperformance which is greater than 5%. The test for overidentifying restriction in Figure 8.a shows no distinct differences from the asymptotic distribution, and the findings are the same as with fixed-weight strategies. Notice that the use of the asymptotic critical values would cause an overrejection of neutral performance (as was shown in Figure 4), whereas the empirical critical values overcome this.

In Section 2.3 we argued that it is important to tie the mean of the SDF to a reasonable value. Otherwise we may evaluate at a misleading tangency

---

[15]The empirical densities shown in this and subsequent figures are estimated with a Gaussian kernel were the bandwidth parameter is chosen automatically (see (Silverman 1986)).

point. We therefore investigate how sensitive the results are to deviations from $\mathrm{E}m_t = \mathrm{E}(1/R_{ft})$. Adding $\pm 1\%$ (on an annual basis) to $\mathrm{E}m_t$ changes the distributions of the performance measures (both fixed-weight and dynamic m_t) fairly little. Adding (subtracting) 1% moves the distributions somewhat to the left (right), yielding a somewhat more negative (positive) performance. The deviations are not enough though to overturn the results.

Up to now we have analyzed unconditional (average) performance. It is also of interest to test the *conditional performance* of the funds, again using *dynamic benchmark strategies*. We condition the unconditional performance measure on the state of the economy to study if the funds have superior/inferior performance in some (predictable) states of the economy. Notice, that the specification of the SDF has to be changed in order to price these conditional strategies under the null (see Section 2.2). In Figure 9.a the test statistics and the empirical distribution of the test statistics are depicted (it now correspond asymptotically to a $\chi^2(4)$-distributed variable).[16] The non-neutral performance appears this time to be more significant for a large portion of the funds. The null is rejected though we use the empirical critical values which are significantly larger than the asymptotic ones. In Figure 9.b, we show the performance measure which correspond to the constant in z_{t-1}. Recall that the performance measures related to the non-constant elements in z_{t-1} (the z_{t-1}^c:s) can be written as

$$\mathrm{E}\lambda_{2t}^c = \mathrm{Cov}(\lambda_{2t}^{uc}, z_{t-1}^c) + \mathrm{E}\lambda_{2t}^{uc}\mathrm{E}z_{t-1}^c. \tag{17}$$

Since $\mathrm{E}z_{t-1}^c$ is set to one, the $\mathrm{E}\lambda_{2t}^c$ measures the sum of the covariance of the unconditional performance and the information variables, and the unconditional performance. In Figures 9.c-e, we show the covariances in (17), that is, the $\mathrm{Cov}(\lambda_{2t}^{uc}, z_{t-1}^c)$:s. The average unconditional performance across the funds (the average of the $\mathrm{E}\lambda_{2t}^{uc}$:s) is slightly higher than before (about 0.9% per year). The unconditional performance measure seems to be positively related to two of the information variables – the general market and the level of the yield curve – as measured by the covariances. Hence, the tendency for non-neutral performance is strengthened (they show a performance of 1.1%-1.2% on average with medians of about 1.4%). That is, when we have experienced an increase in the stock market or a high level of interest rates in general, the funds have performed better than on average. The opposite

---

[16]In the conditional evaluation, one fund is not examined due to numerical difficulties in forming the test statistic.

is the case for the bond excess return instrument. Overall, the stronger case for non-neutral performance appears to stem from the fact that the unconditional performance can be predicted by the information variables.

To sum up, we present the means and medians of the performance measures for the three cases in Table 5. Decompositions of the performance measures are also given. Using (6), the performance measures can be written as

$$\mathrm{E}\lambda_{2t}/\mathrm{E}m_t = [\mathrm{E}\left(z_{t-1} \otimes R_{2t}\right) - \mathbf{1}/\mathrm{E}m_t] - [-\mathrm{Cov}\left(z_{t-1} \otimes R_{2t}, m_t\right)/\mathrm{E}m_t], \quad (18)$$

where we have used the fact that the instrument are normalized such that $\mathrm{E}z_{t-1} = \mathbf{1}$. The first term within brackets can be interpreted as the managed returns of the funds in excess over the implied risk-free rate. The second term within brackets is then the 'risk-premium.' From the table we see that the average excess return over a risk-free rate in the unconditional evaluation is about 3.4%, and that the risk premium is about 2.9%, yielding a superior performance of about 0.5% per year. Noteworthy is that in the conditional evaluation, the excess returns (for the different information variables) are about the same magnitude or higher, whereas the risk premia are smaller.

An assessment of the various stochastic discount factors used, reveal that their means are similar (since they are tied to the bill portfolio, the only differences come from the positivity imposed). On average, the SDFs imply a risk-free rate of about 9% per year. It is evident that when strategies based on the information variables are allowed, the standard deviations of the SDFs increase significantly. The dynamic SDFs show a correlation about 56%, whereas their correlations to the fixed-weight SDF are fairly low (about 20%). This indicates that risk premia would be very different if the 'correct' specifications of the SDF are not being used.

For six of the funds we have data from 1987. Applying the above tests on the funds over this longer sample period shows that for 5 of the 6 funds the performance have been worse during the period 1987-95 compared to the period 1991-95. This holds for both the unconditional and the conditional setting. Two funds change from positive to negative performance as the sample is extended backwards.

Finally, we assess the robustness of the results by undertaking the evaluation using size portfolios instead of industry portfolios as benchmarks. This changes the unconditional performance measures very little. The distributions are centered around zero for a fixed-weight SDF, the evaluation using

23

a dynamic SDF indicate better performance, but the null of neutral performance cannot be rejected.

6 Conclusions

In this paper we provide evidence on the use of stochastic discount factors in the evaluation of portfolio performance. The paper is divided into three major parts. In the first part we build on the papers by (Bansal and Harvey 1996), (Chen and Knez 1996), and (Farnsworth, Ferson, Jackson, and Todd 1996) and discuss performance evaluation using efficient (by construction) benchmarks and allowing for dynamic strategies. We relate the performance measures in (Chen and Knez 1996) to the traditional mean-variance analysis. Especially, we make an explicit interpretation of them in terms of a specific Jensen's alpha measures computed against efficient benchmarks. Moreover, we highlight the importance of getting the mean of the SDF right. Otherwise, the evaluation could be undertaken at an arbitrary (and misleading) point in the mean-variance space. Finally, we show how to select the instruments in order to undertake a correct conditional evaluation, which allows for dynamic strategies. If we fail to use the correct instruments in constructing the SDF, then a fund with a neutral dynamic trading strategy can incorrectly show either inferior or superior performance.

In the second part of the paper we examine the small sample properties of GMM estimators in this setting by a series of Monte Carlo simulations. The results provide no evidence against a prespecified weighting matrix. We find evidence in favor of using weekly non-overlapping returns instead of monthly returns. We also show that the size can sometimes be seriously distorted such that empirical critical values should be used. However, when we correct for the size, the power is about the same in the different settings. Moreover, we highlight the case that an economically significant excess return is needed in order to reject the null of neutral performance, unless the sample is very long. The power properties are better in the unconditional setting compared to the conditional setting and/or when the econometric systems are large. Moreover, the intersection test appears to have higher power than the spanning test.

In the third part, we apply the methodology to Swedish-based mutual funds and offer an actual evaluation. An unconditional as well as a conditional evaluation is undertaken. In the unconditional evaluation, using either

fixed-weight or dynamic strategies (via returns scaled by available public information), we cannot find strong evidence of non-neutral performance. The performance of the funds is on average positive, but it cannot be confirmed statistically. However, when we evaluate conditionally, again allowing for dynamic strategies, the results indicate a tendency towards non-neutral performance. This captures the degree of predictability in the performance measure.

A   Simple Benchmark Strategies and Performance Measures

Let R_{1t} be an $n_1 \times 1$ vector, x_t an $k \times 1$ vector, and z_t an $l \times 1$ vector. Take expectations of (7)

$$\mathrm{E}\lambda_{1t} = \mathrm{E}\left(z_{t-1} \otimes x_{t-1} \otimes R_{1t}\right) m_t - \mathrm{E}z_{t-1} \otimes x_{t-1} \otimes \mathbf{1}_{n_1}, \tag{19}$$

which is set to zero by the choice of m_t. (The case with a fixed-weight benchmark portfolio is obtained by setting $x_t = 1$.) This appendix shows that any portfolio with weights which are linear in the information variables x_{t-1}, and summing to one in every period, will then be assigned a neutral performance.

Combine (1) and (2) to get

$$R_t = x'_{t-1}\delta R_{1t}, \text{ such that } x'_{t-1}\delta \mathbf{1}_{n_1} = 1, \tag{20}$$

where δ is a $k \times n_1$ matrix.

For simplicity, suppose x_t has only two elements: a constant and a random variable, $x'_t = \begin{bmatrix} 1 & \tilde{x}_t \end{bmatrix}$. The portfolio weights are then

$$x'_t\delta = \begin{bmatrix} 1 & \tilde{x}_t \end{bmatrix} \begin{bmatrix} \delta_{11} & \cdots & \delta_{1n_1} \\ \delta_{21} & \cdots & \delta_{2n_1} \end{bmatrix}. \tag{21}$$

The portfolio weights must sum to one in every period, that is,

$$\begin{aligned} 1 &= x'_t\delta\mathbf{1}_{n_1} \\ &= \begin{bmatrix} 1 & \tilde{x}_t \end{bmatrix} \begin{bmatrix} \delta_{11} + \cdots + \delta_{1n_1} \\ \delta_{21} + \cdots + \delta_{2n_1} \end{bmatrix}. \end{aligned} \tag{22}$$

The only way to guarantee that this holds in every period is to have the first row in δ summing to one and the second row summing to zero. This means that movements in the stochastic information variables are only allowed to change the portfolio weights in a zero investment portfolio of R_{1t}. More formally, we have the requirement that

$$\delta\mathbf{1}_{n_1} = \begin{bmatrix} 1 \\ \mathbf{0} \end{bmatrix}. \tag{23}$$

The $l \times 1$ vector of performance measures of the scalar benchmark portfolio R_t in (20) is (one for each instrument in z_{t-1})

$$\mathrm{E}\lambda_t = \mathrm{E}\left(z_{t-1}x'_{t-1}\delta R_{1t}\right)m_t - \mathrm{E}z_{t-1}. \tag{24}$$

Pick out the elements in $\mathrm{E}\lambda_{1t}$ from (19) and $\mathrm{E}\lambda_t$ from (24) which correspond to the jth element in z_{t-1}, $z_{t-1,j}$, and write them as

$$\mathrm{E}\lambda_{1t,j} = \mathrm{E}\left(z_{t-1,j}x_{t-1} \otimes R_{1t}\right)m_t - \mathrm{E}z_{t-1,j}x_{t-1} \otimes \mathbf{1}_{n_1}, \text{ and} \tag{25}$$

$$\mathrm{E}\lambda_{t,j} = \mathrm{E}\left(z_{t-1,j}x'_{t-1}\delta R_{1t}\right)m_t - \mathrm{E}z_{t-1,j}. \tag{26}$$

Let δ_i be the ith row of δ, so

$$x'_{t-1}\delta R_{1t} = \delta_1 R_{1t} + \tilde{x}_{t-1}\delta_2 R_{1t}. \tag{27}$$

This allows us to rewrite (26) as

$$\mathrm{E}\lambda_{t,j} = \delta_1\mathrm{E}\left(z_{t-1,j}R_{1t}m_t\right) + \delta_2\mathrm{E}\left(z_{t-1,j}\tilde{x}_{t-1}R_{1t}m_t\right) - \mathrm{E}z_{t-1,j}. \tag{28}$$

Since the first element in x_{t-1} is a constant, (25) shows that each of the elements in the vector $\mathrm{E}(z_{t-1,j}R_{1t}m_t)$ equals $\mathrm{E}z_{t-1,j}$. From (23) we also know that $\delta_1 \mathbf{1} = 1$. Combining these facts gives that $\delta_1\mathrm{E}(z_{t-1,j}R_{1t}m_t)$ in (28) equals $\mathrm{E}z_{t-1,j}$.

Similarly, by (25), each of the elements in $\mathrm{E}(z_{t-1,j}\tilde{x}_{t-1}R_{1t}m_t)$ equals $\mathrm{E}(z_{t-1,j}\tilde{x}_{t-1})$. Since $\delta_2 \mathbf{1} = 0$, we get that $\delta_2\mathrm{E}(z_{t-1,j}\tilde{x}_{t-1}R_{1t}m_t)$ in (28) is zero. Together, these results give $\mathrm{E}\lambda_{t,j} = 0$ in (28). To sum up, any portfolio with weights which are linear in the information variables x_{t-1} and summing to one in every period, has a neutral performance if $\mathrm{E}\lambda_{1t} = 0$ in (19).

B   GMM Estimation and Testing

In this section we briefly review the limiting distributions of the parameter estimate and the test for overidentifying restrictions.

Let Λ_t be the stacked moment conditions in (12), and let Q denote the Jacobian of the sample moment conditions, that is,

$$Q = \frac{\partial\bar{\Lambda}}{\partial\gamma'}, \tag{29}$$

where $\bar{\Lambda} = \frac{1}{T}\sum_{t=1}^{T}\Lambda_t$. The general (for an arbitrary weighting matrix) expression for the asymptotic covariance matrix of $\sqrt{T}(\hat{\gamma} - \gamma)$ is equal to

$$V = (Q'WQ)^{-1}Q'WSWQ(Q'WQ)^{-1}, \tag{30}$$

where

$$S = \text{Cov}(\sqrt{T}\bar{\Lambda}) = \lim_{T\to\infty}\sum_{s=-(T-1)}^{T-1}\text{Cov}(\Lambda_t, \Lambda_{t-s}). \tag{31}$$

It is common to set $W = S^{-1}$, and (30) then simplifies to $V = (Q'S^{-1}Q)^{-1}$. Regarding the overidentifying restrictions, (Hansen 1982) shows (Lemma 4) that the asymptotic covariance matrix of the moment restrictions $\sqrt{T}(\bar{\Lambda} - \Lambda)$ is

$$\Omega = [I - Q(Q'WQ)^{-1}Q'W]S[I - WQ(Q'WQ)^{-1}Q'] \tag{32}$$

When $W = S^{-1}$, this simplifies to $\Omega = S - Q(Q'S^{-1}Q)^{-1}Q'$ which is numerically equal to S (see (Cochrane 1997)). In either case, Ω has reduced rank equal to the number of overidentifying restrictions, denoted q below. We can therefore construct a χ^2-distributed test statistic

$$T\bar{\Lambda}'\Omega^+\bar{\Lambda} \sim \chi^2(q), \tag{33}$$

where $^+$ indicates a generalized inverse (see (Cochrane 1996)). It is tricky to calculate Ω^+ with reasonable accuracy. According to our experience, it is very useful to impose the known rank of Ω in these calculations. This is done in several steps. First, calculate the singular value decomposition such that $usv' = \Omega$, reorder u, s, and v so that the singular values come in decreasing order. Second, partition $s = \begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix}$, $u = \begin{bmatrix} u_1 & u_2 \end{bmatrix}$, and $v = \begin{bmatrix} v_1 & v_2 \end{bmatrix}$ so that the first block contains the q largest singular values (the remaining singular values should, in principle, be zero). Third, the generalized inverse is calculated as $\Omega^+ = v_1 s_1^{-1} u_1'$.

C   The Data Generating Processes

This appendix describes the data generating processes for the reference assets and the portfolios under evaluation in the Monte Carlo simulations. We want a parsimonious modeling for the reference assets, but still capturing

time-variation in the first and second moments. We therefore use a first-order vector autoregressive model (VAR) for the first moments. The dynamics of the second moments is a simple version of the general BEKK model in (Engle and Kroner 1995). The advantage of the BEKK model for conditional covariances is that it guarantees positive definite conditional covariance matrices under weak conditions. Moreover, compared to other models it uses few parameters, but still allows for conditional correlations and it is able to capture potential cross-volatility interactions as well. For tractability, the model is here restricted to be of order (1,1). The specification below allows us to simultaneously model conditional means, variances, covariances, and correlations.

Let $\tilde{z}_t$ be an $k-1$ vector of instruments. The dynamics for the $\ln \tilde{z}_t$ is a VAR(1) with BEKK errors

$$\ln \tilde{z}_t = \mu + A \ln \tilde{z}_{t-1} + \epsilon_t, \tag{34}$$

where μ is an $k-1$ parameter vector, A is an $(k-1) \times (k-1)$ parameter matrix, and where ϵ_t is a $k-1$ vector of error terms which are assumed to be conditionally multivariate normal distributed, but serially uncorrelated. That is,

$$\epsilon_t \mid \Omega_{t-1} \sim \mathrm{N}(0, H_t), \tag{35}$$

where Ω_{t-1} denotes the information set at time $t-1$. Assuming that the dynamics of the conditional covariance matrix is covariance stationary with an unconditional covariance matrix equal to H_0, we estimate a diagonal BEKK model that can be written as

$$H_t = H_0 \odot (\mathbf{1}_n \mathbf{1}_n' - bb' - cc') + bb' \odot \epsilon_{t-1}\epsilon_{t-1}' + cc' \odot H_{t-1}, \tag{36}$$

where $\odot$ denotes the Hamadard product (element-by-element), and b and c are parameter vectors.

The parameters are estimated in a two-step procedure. In the first step we estimate the VAR system (using Least Squares) and the parameters in H_0. In the second step the BEKK parameters, b and c, are estimated by Maximum Likelihood.

The dynamics of the benchmark portfolios are

$$\ln R_{1t} = C \begin{bmatrix} 1 \\ \ln \tilde{z}_{t-1} \end{bmatrix} + \eta_t, \tag{37}$$

where C is an $n_1 \times k$ parameter matrix, and η_t a vector of iid normally distributed error terms with covariance matrix Σ. C and Σ are estimated using Least Squares.

The data generating process of the portfolios that are being evaluated under the null hypothesis is

$$R_{2t} = BR_{1t} + u_t, \tag{38}$$

where $B\mathbf{1} = \mathbf{1}$ and u_t is an vector of iid normally distributed errors with covariance matrix Ψ. Ψ is taken from a Least Squares estimation of (38) of a typical mutual funds, and B is also based on the estimate, but adjusted to satisfy the condition $B\mathbf{1} = \mathbf{1}$ exactly. Note that (38) is actually the case of spanning in (Huberman and Kandel 1987). The advantage of this approach is that the funds are generated only by the benchmarks, and not at all by the zero-beta asset, which is important for the conditional evaluation. In the actual evaluation we still use the intersection test. Under the alternative hypothesis of superior performance we add a constant yielding, for instance, a 5% excess return on an annual basis.

D   Saturation Ratios

In GMM estimation, the total number of observations divided by the number of parameters to be estimated (including the number of unique parameters in the weighting matrix) is sometimes referred to as the saturation ratio. The total number of observations is equal to the number of moment conditions times the length of the data. The saturation ratios for the three systems we consider in this paper are given by

$$SR^1 = \frac{(n_1 + n_2 + 1)\, T}{\left[n_1 + 1 + \frac{(n_1+n_2+1)(n_1+n_2+2)}{2} \right]}, \tag{39}$$

$$SR^2 = \frac{(kn_1 + n_2 + 1)\, T}{\left[kn_1 + 1 + \frac{(kn_1+n_2+1)(kn_1+n_2+2)}{2} \right]}, \tag{40}$$

and

$$SR^3 = \frac{\left(\frac{k(k+1)}{2} n_1 + kn_2 + 1 \right) T}{\left[\frac{k(k+1)}{2} n_1 + 1 + \frac{\left(\frac{k(k+1)}{2} n_1 + kn_2 + 1 \right)\left(\frac{k(k+1)}{2} n_1 + kn_2 + 2 \right)}{2} \right]}, \tag{41}$$

30

where SR^1 refer to the fixed-weight/unconditional case, SR^2 to the dynamic/unconditional case, and SR^3 to the dynamic/conditional case.

References

Andrews, Donald W. K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* 59, 817–858.

Bansal, Ravi, and Campbell R. Harvey, 1996, Performance evaluation in the presence of dynamic trading strategies, Working Paper, Duke University.

Bansal, Ravi, and S. Viswanathan, 1993, No arbitrage and arbitrage pricing: a new approach, *Journal of Finance* 48, 1231–1262.

Bekaert, Geert, and Michael S. Urias, 1996, Diversification, integration and emerging market closed-end funds, *Journal of Finance* 51, 835–869.

Chamberlain, Gary, and Michael Rothschild, 1983, Arbitrage, factor structure, and mean-variance analysis on large asset markets, *Econometrica* 51, 1281–1304.

Chen, Zhiwu, and Peter J. Knez, 1996, Portfolio measurement: theory and applications, *Review of Financial Studies* 9, 511–555.

Cochrane, John H., 1996, A cross-sectional test of an investment-based asset pricing model, *Journal of Political Economy* 104, 572–621.

———, 1997, Asset pricing, Manuscript, University of Chicago.

Dahlquist, Magnus, and Paul Söderlind, 1997, Swedish buy-and-hold stock portfolios: 1985 to 1995, Working Paper, Stockholm School of Economics.

De Santis, Giorgio, 1995, Volatility bounds for stochastic discount factors: tests and implications from international financial markets, Working Paper, University of Southern California.

Dybvig, Philip H., and Stephen A. Ross, 1985, The analytics of performance measurement using a security market line, *Journal of Finance* 40, 401–416.

Engle, Robert F., and Kenneth F. Kroner, 1995, Multivariate simultaneous Generalized ARCH, *Econometric Theory* 11, 122–150.

Farnsworth, Heber, Wayne Ferson, David Jackson, and Steven Todd, 1996, Conditional performance evaluation, Working Paper,University of Washington.

Ferson, Wayne E., 1995, Theory and empirical testing of asset pricing models, in Robert A. Jarrow, Vojislav Maksimovic, and William T. Ziemba, ed.: *Handbooks in Operations Research and Management Science* . pp. 145–200 (North-Holland, Amsterdam).

———— , and Stephen R. Foerster, 1994, Finite sample properties of the generalized method of moments in tests of conditional asset pricing models, *Journal of Financial Economics* 36, 29–55.

Ferson, Wayne E., and Rudi Schadt, 1996, Measuring fund strategy and performance in changing economic conditions, *Journal of Finance* 51, 425–461.

Gallant, Ronald A., and George Tauchen, 1989, Seminonparametric estimation of conditionally constrained heterogeneous processes: asset pricing implications, *Econometrica* 57, 1091–1120.

Glosten, Lawrence, and Ravi Jagannathan, 1994, A contingent claims approach to performance evaluation, *Journal of Empirical Finance* 1, 133–166.

Hansen, Lars Peter, 1982, Large sample properties of Generalized Method of Moments estimators, *Econometrica* 50, 1029–1054.

———— , and Ravi Jagannathan, 1991, Implications of security market data for models of dynamic economies, *Journal of Political Economy* 99, 225–262.

———— , 1997, Assessing specification errors in stochastic discount factor models, *Journal of Finance* 52, 557–590.

Hansen, Lars Peter, and Scott F. Richard, 1987, The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models, *Econometrica* 55, 587–613.

Huberman, Gur, and Shmuel Kandel, 1987, Mean-variance spanning, *Journal of Finance* 42, 873–888.

Newey, Whitney K., and Kenneth D. West, 1987, A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

———— , 1994, Automatic lag selection in covariance matrix estimation, *Review of Economic Studies* 61, 631–654.

Richardson, Matthew, and Tom Smith, 1994, A unified approach to testing for serial correlation on stock returns, *Journal of Business* 67, 371–399.

Roll, Richard, 1978, Ambiguity when performance is measured by the security market line, *Journal of Finance* 38, 1051–1069.

Silverman, Bernhard W., 1986, *Density estimation for statistics and data analysis* (Chapman and Hall, London).

Snow, Karl N., 1991, Diagnosing asset pricing models using the distribution of asset returns, *Journal of Finance* 46, 955–983.

Figure 1: The Relation to Mean-Standard Deviation Frontiers.

The figures show the relation between the SDF tests (without positivity imposed) and mean-standard deviation frontiers. Figure a shows the mean-standard deviation frontier of the benchmark assets R_{1t} (solid hyperbola), and the frontier of the benchmarks plus a fund, R_{2t}, with neutral performance (dashed hyperbola). The frontiers intersect at the tangency portfolio R_{Tt}^*, marked by a circle. The return on the SDF, R_{bt}^*, is marked by a square. The implied risk free rate from the SDF is $1/\mathrm{E}m_t^*$. Figure b shows the frontiers when the fund has an excess return of 5%. The frontiers still intersect but now at the inefficient part, marked by a triangle. This tangency portfolio, R_{Tt}^0, corresponds to a risk free rate of R^0.

a. Neutral performance

b. 5% Excess return
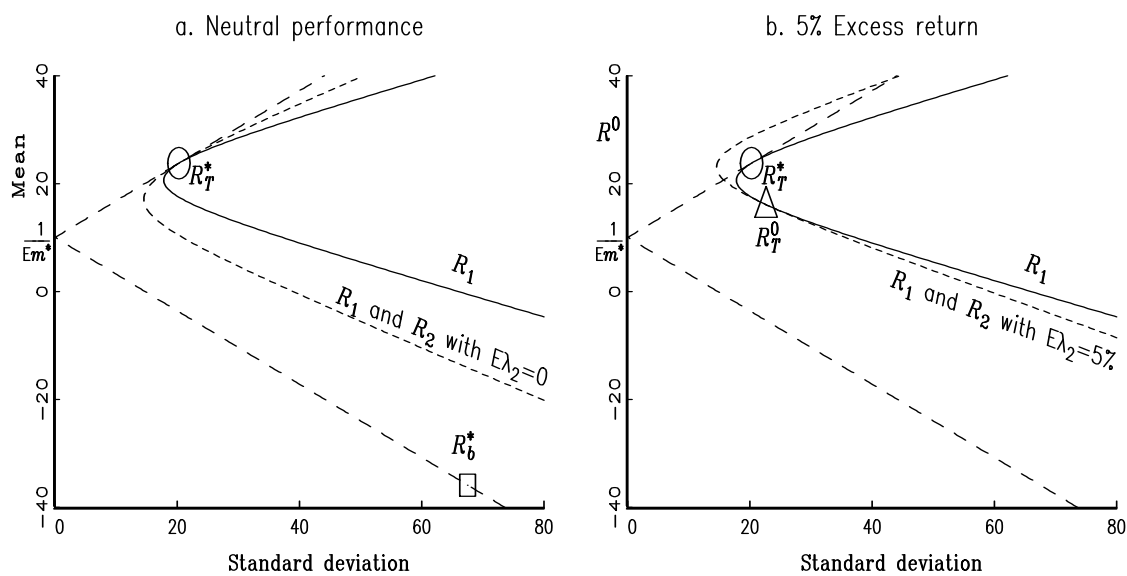
Table 1:   Distributional Characteristics of Benchmarks and Information Variables.

The table shows moments for the information variables and the weekly gross returns of the benchmark portfolios. The benchmark portfolios are passive buy-and-hold industry portfolios. The short-term bond return is a holding-return on a 6-month Treasury yield. The information variables are the following. First, the annualized yield on the 6-month Treasury detrended with a lagged two-month moving average. Second, the return on a long-term bond (maturity between 3-5 years) in excess of the return on the short-term bond. Third, the gross return on a passive buy-and-hold portfolio of the general stock market. The mean and standard deviations of returns are expressed as net returns and in percent per year. The kurtosis is in excess of the kurtosis for a normally distributed variable (3).

	Mean	Standard Deviation	Skewness	Excess Kurtosis	Auto-correlation
Benchmarks					
Manufacturing	17.43	23.25	-0.03	3.31	0.12
Wood and pulp	16.65	29.83	0.95	9.25	0.11
Retail and construction	13.64	29.29	0.97	6.91	0.14
Bank and insurance	15.39	27.60	0.35	4.13	0.15
Miscellaneous	20.57	18.07	-0.46	2.94	0.11
Short-term bond	10.04	1.29	-0.08	24.82	-0.10
Information variables					
Yield curve level	0.02	0.15	1.65	8.90	0.87
Bond excess return	1.70	5.46	-0.26	9.02	-0.11
General stock index	16.74	20.40	-0.25	3.67	0.16

Figure 2: Empirical Size and Power: Choice of Weighting Matrix.

The figures show the size and power for five different weighting matrices (see be-
low). The results are for the test of unconditional performance with dynamic benchmarks,
on a sample of 250 weekly observations. Figure a shows the empirical size versus nominal
size, and Figure b shows the corresponding power, using empirical critical values from
Figure a, when the fund has 5% excess return. The first estimator is a two-step GMM.
It uses an identity matrix as weighting matrix in the first step, and the inverse of Newey
and West (1987) covariance estimator with one lag in the second step. The second is
similar to the first, but iterates until the change in loss function value is less than 1%
of the expected value under the null hypothesis (here, one). The third and fourth are
similar to the first, but use the covariance estimator with automatic lag order selection
suggested by Newey and West (1994) and Andrews (1991), respectively. The fifth is a
one-step estimator using the inverse of the second moments of the managed portfolios
and the fund return.

a. Size

b. Power at 5% excess return

--- Newey–West –87, 2 steps
········· Newey–West –87, iterated
--- Newey–West –94, 2 steps
—— Andrews –91, 2 steps
—— Prespecified, $W=(ERR')^{-1}$

37
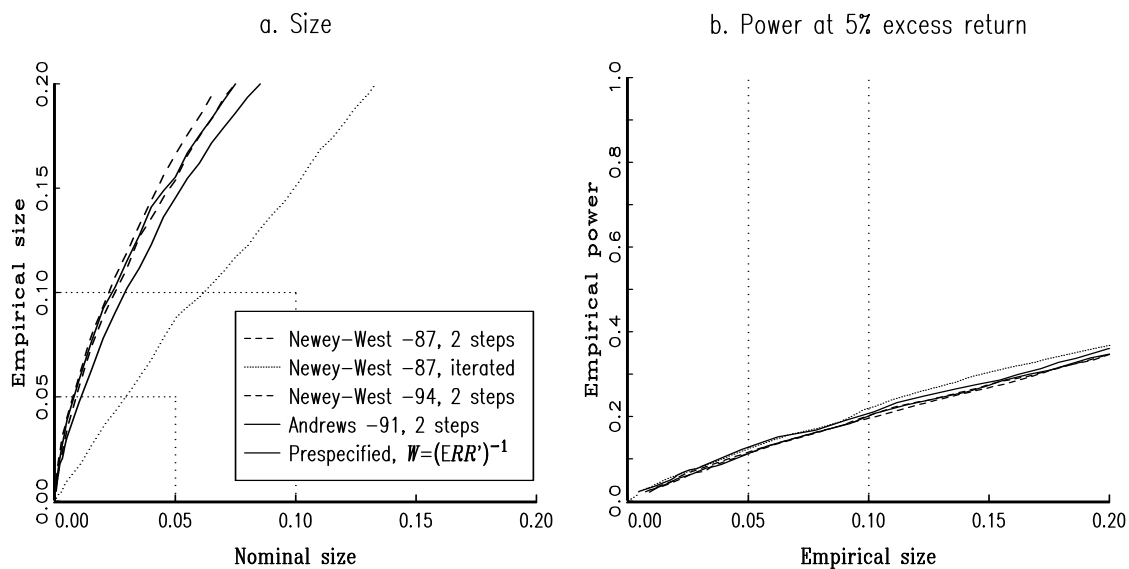
Figure 3: Empirical Size and Power: Different Performance Tests.

The figures show the size and power of the three performance tests used in the paper: unconditional performance with fixed-weight or dynamic benchmarks, and conditional performance with dynamic benchmarks. Figure a shows the empirical size versus nominal size. The sample has 250 weekly observations, and a prespecified weighting matrix is used. Figure b shows the corresponding power, using empirical critical values from Figure a, when the fund has 5% excess return.


a. Size   b. Power at 5% excess return
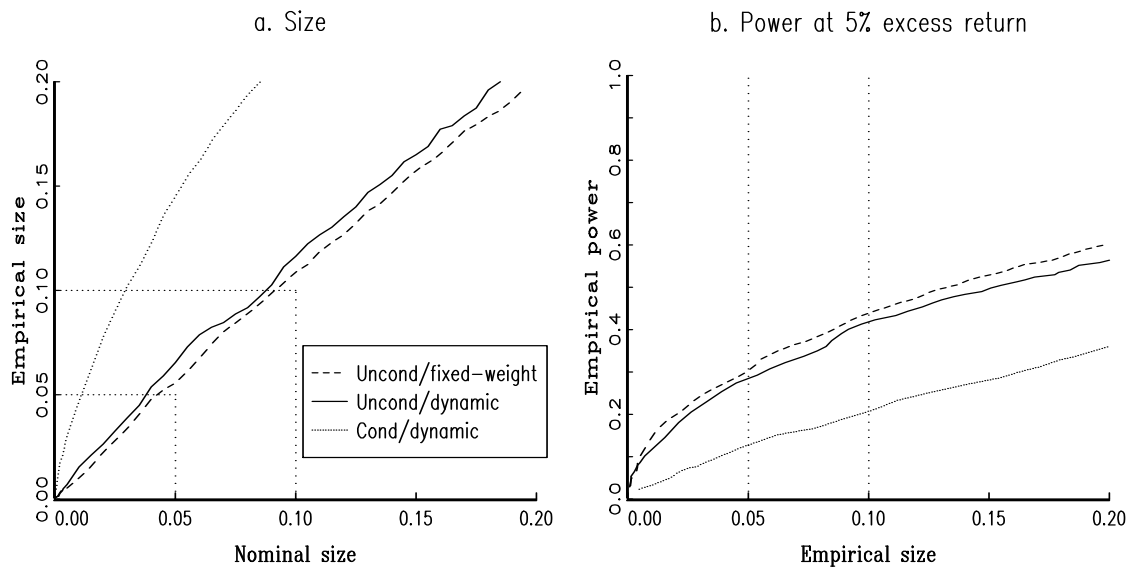
Figure 4: Empirical Size and Power: Sample Length and Excess Return.

The figures show the size and power for two different sample sizes and excess returns. The results are for the test of unconditional performance with dynamic benchmarks, using a prespecified weighting matrix. Figure a shows the empirical size versus nominal size for samples of 250 and 500 weekly observations. Figure b shows the corresponding power, using empirical critical values from Figure a, when the fund has 3% or 5% excess return.
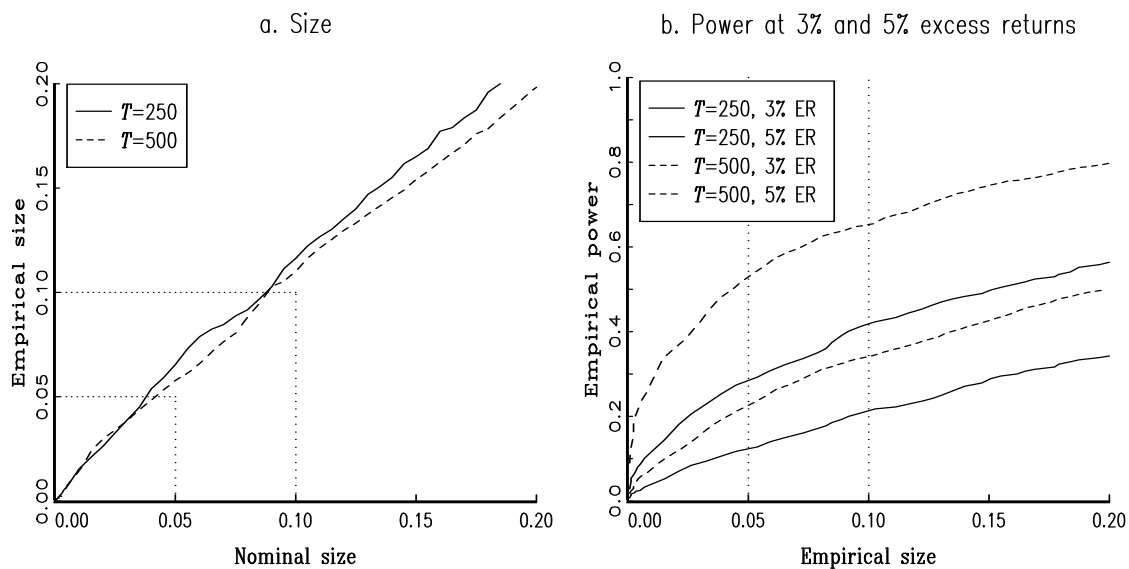
Figure 5: Empirical Size and Power: Evaluation Horizon and Sampling Frequency.

The figures show the size and power for three different types of data sampling from a sample of 248 weeks: weekly returns, weekly sampled monthly returns, and monthly sampled monthly returns. The results are for the test of unconditional performance with dynamic benchmarks, using a prespecified weighting matrix. Figure a shows the empirical size versus nominal size, and Figure b shows the corresponding power, using empirical critical values from Figure a, when the fund has 5% excess return.
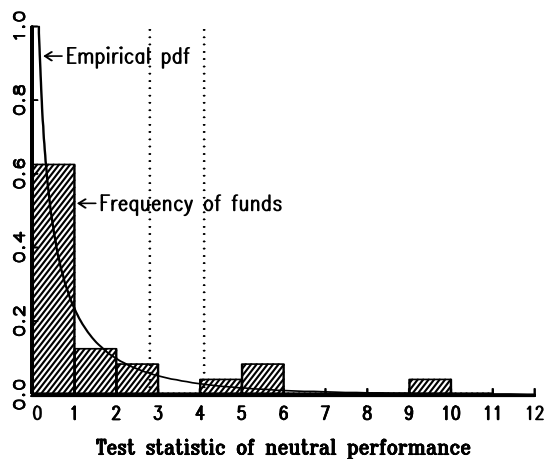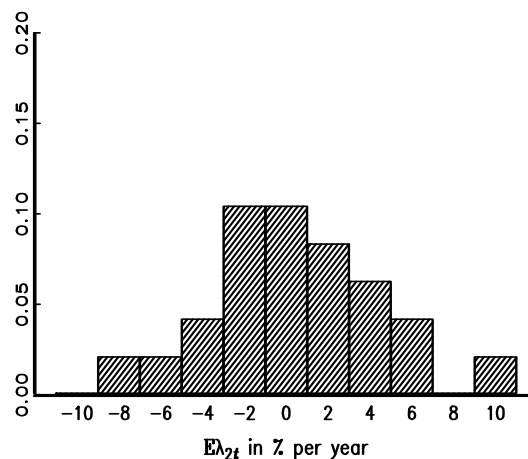

a. Size

b. Power at 5% excess return
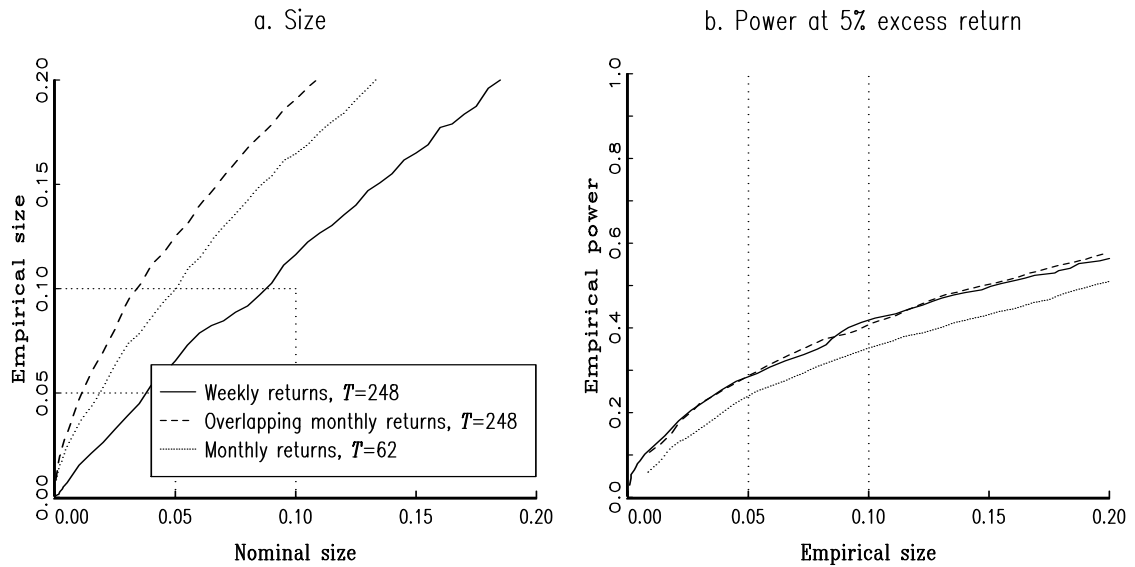
Figure 6: Empirical Size and Power: Intersection versus Spanning.

The figures show the size and power of intersection and spanning tests of unconditional performance with dynamic benchmarks. A prespecified weighting matrix is used on a sample with 250 weekly observations. Figure a shows the empirical size versus nominal size, and Figure b shows the corresponding power, using empirical critical values from Figure a, when the fund has 5% excess return.

a. Size

b. Power at 5% excess return
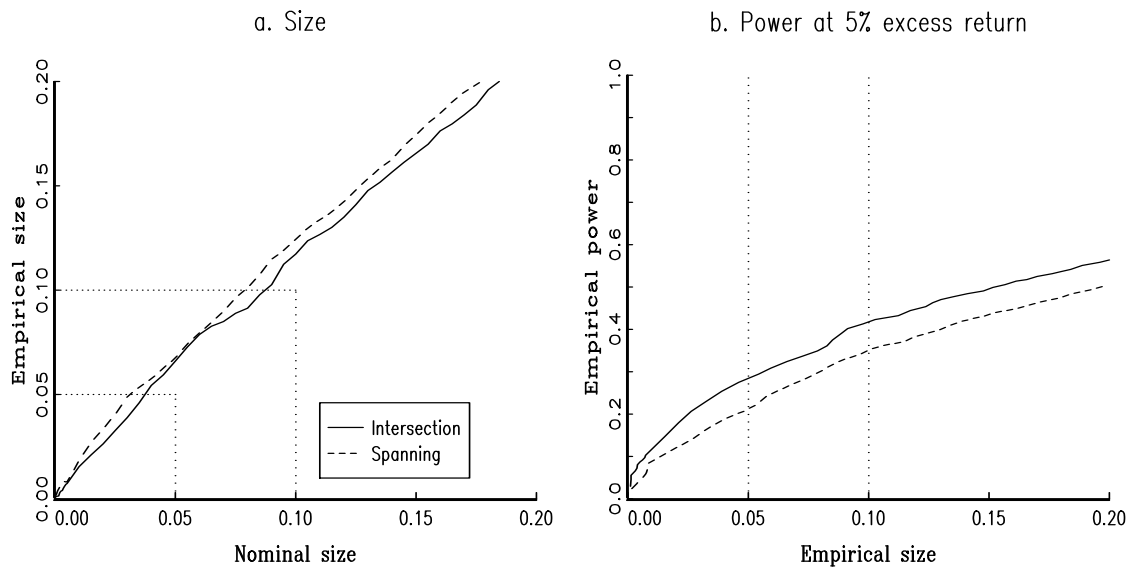
Figure 7: Unconditional Performance using a Fixed-Weight SDF.

The figures show the result of an unconditional evaluation using a fixed-weight benchmark. Figure a shows the frequency distribution of the test statistic of neutral performance (one overidentifying restriction). The empirical density function (asymptotically corresponding to a $\chi^2(1)$-distribution under the null hypothesis) is also depicted. The vertical dotted lines mark the empirical critical values at the 5% and 10% significance level. Figure b shows the distribution of performance measures, $E\lambda_{2t}$, which are based on a positive SDF.
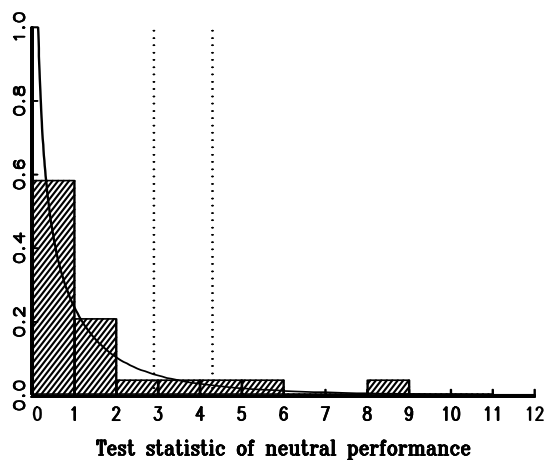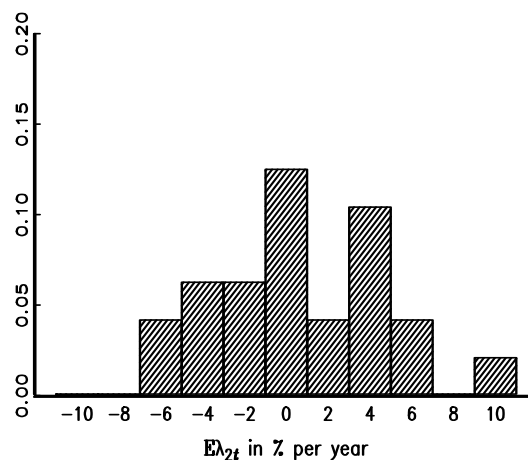
Figure 8: Unconditional Performance using a Dynamic SDF.

The figures show the result of an unconditional evaluation using a dynamic bench-mark. Figure a shows the frequency distribution of the test statistic of neutral performance (one overidentifying restrictions). The empirical density function (asymp-totically corresponding to a $\chi^2(1)$-distribution under the null hypothesis) is also depicted. The vertical dotted lines mark the empirical critical values at the 5% and 10% significance level. Figure b shows the distribution of performance measures, $E\lambda_{2t}$, which are based on a positive SDF.

a. Test statistic

b. Performance measure

Test statistic of neutral performance
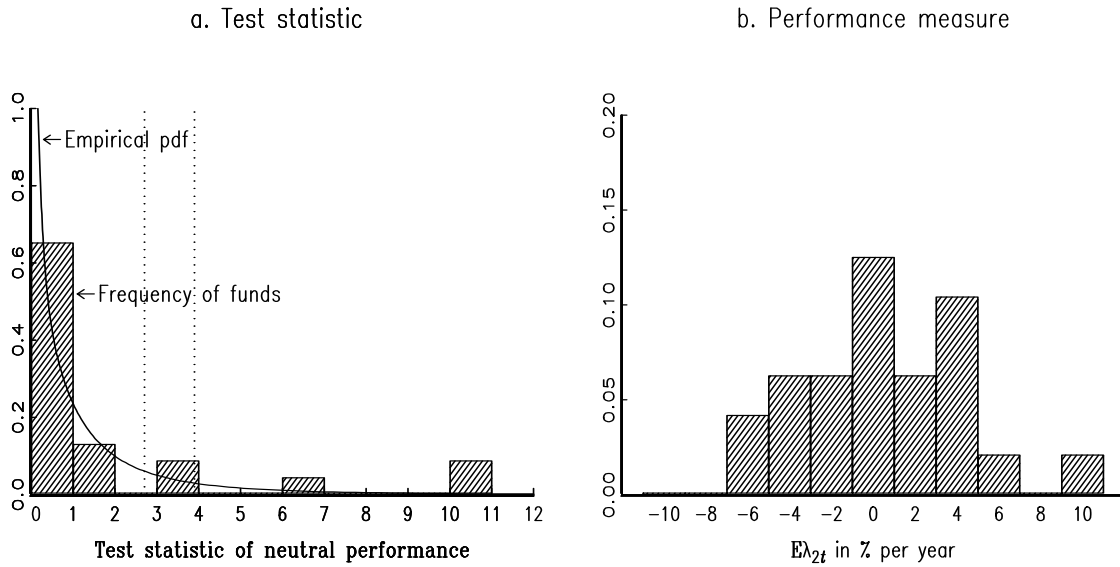
$E\lambda_{2t}$ in % per year

43

Figure 9: Conditional Performance using a Dynamic SDF.

The figures show the result of a conditional evaluation using a dynamic benchmark. Figure a shows the frequency distribution of the test statitic of neutral performance (four overidentifying restrictions). The empirical density function together with the asymptotically corresponding density under the null hypothesis (a $\chi^2(4)$-distribution)) are also depicted. The vertical dotted lines mark the empirical critical values at the 5% and 10% significance level. Figure b shows the distribution of the performance measures which correspond to the constant element in z_{t-1}. Figures c-e show the distributions of the covariances of the unconditional performance measures and the predetermined instruments, $\text{Cov}(\lambda_{2t}^{uc}, z_{t-1}^c)$. That is, the covariances shown are assocciated with the general market, the yield curve level, and the return on a long-term bond in excess of a short-term bond.
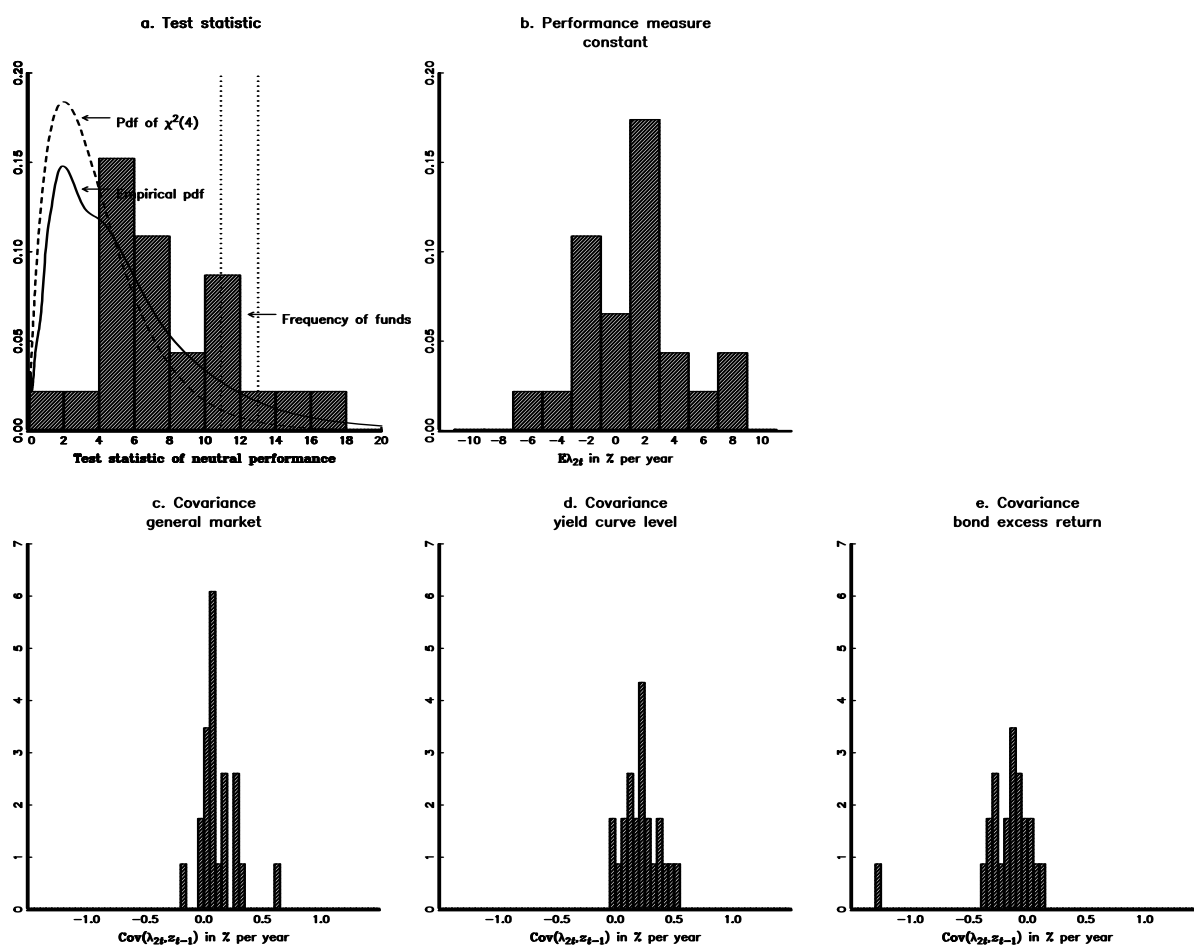

44

Table 2: A Summary of the Performance Evaluations.

The table summarizes the results of the three evaluations which are undertaken on a sample of Swedish mutual funds. The first evaluation is unconditional and utilizes a fixed-weight benchmark. The second evaluation is also unconditional, but allows for dynamic strategies through predetermined information variables. The variables (in addition to a constant) are the general market return, the level of the yield curve (detrended) and the return on a long-term bond in excess of the return on the short-term bond. The third evaluation is conditional on the state of the economy, captured by the above information variables. Again, dynamic strategies are allowed. The conditional evaluation yields a several-dimensional performance measure – one per information variable. The first two rows in the table show the means and medians (in % per year) across the funds. In the subsequent rows the average performance is decomposed into an excess return and a risk premia as expressed in (5.2). All measures are in % per year.

	Unconditional Fixed-weight Const.	Unconditional Dynamic Const.	Conditional Dynamic Const.	General	Yield	Bond
Performance measures						
Mean	0.48	0.42	0.90	1.12	1.18	0.73
Median	0.08	0.14	1.24	1.42	1.40	0.78
Decomposition						
Excess return	3.43	3.16	3.43	3.97	3.72	4.15
Risk premium	2.95	2.74	2.53	2.85	2.54	3.42