Distribution Agreement

Signature:

__________________________ __________________________

Sermin Gungor Date

Market Liquidity and Linear Factor Pricing Models: Empirical Assessment and New Distribution-Free Tests

By

Sermin Gungor
Doctor of Philosophy

Economics

Richard Luger, Ph.D.
Advisor

Esfandiar Maasoumi, Ph.D.
Committee Member

Elena Pesavento, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Market Liquidity and Linear Factor Pricing Models: Empirical Assessment and New Distribution-Free Tests

By

Sermin Gungor

Emory University
Economics Department

Richard Luger, Ph.D
Advisor

Esfandiar Maasoumi, Ph.D
Committee Member

Elena Pesavento, Ph.D
Committee Member

An abstract of
A dissertation submitted to the Faculty of
the James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of Doctor of Philosophy
in Economics
2010

ABSTRACT

Market Liquidity and Linear Factor Pricing Models: Empirical Assessment and New Distribution-Free Tests

By Sermin Gungor

An important problem of modern financial economics is understanding and quantifying the trade-off between risk and expected return. Although we anticipate the riskier assets to yield higher returns, the quantification of the risk-return tradeoff was possible only after the introduction of the linear factor pricing models. Given the crucial role of these models in the asset pricing theory, this dissertation analyzes the linear factor pricing models from both financial economics and econometrics points of view. The first chapter examines the role of time-varying market liquidity in explaining the future asset returns using a conditional multifactor asset pricing framework. The second chapter develops exact distribution-free tests of unconditional mean-variance efficiency. The third chapter proposes a finite-sample procedure to test the beta-pricing representation of linear factor pricing models that is applicable even if the number of test assets is greater than the length of the time series.

Market Liquidity and Linear Factor Pricing Models: Empirical Assessment and New Distribution-Free Tests

By

Sermin Gungor

Emory University
Economics Department

Richard Luger, Ph.D
Advisor

Esfandiar Maasoumi, Ph.D
Committee Member

Elena Pesavento, Ph.D
Committee Member

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of Doctor of Philosophy
in Economics
2010

ACKNOWLEDGMENTS

I would like to express my deepest appreciation for my advisor Dr. Richard Luger for his continuous support throughout the Ph.D. program. He trained me superbly as an economist, taught me how to ask questions, how to approach a problem, and the need to be persistent to accomplish any goal. Most importantly, Richard was always there to listen and give advice. I also would like to thank my committee members Dr. Esfandiar Maasoumi and Dr. Elena Pesavento for their questions, comments, and constant support. Last but not least, I am grateful to everyone in the Economics Department of the Emory University for accepting me as a part of the family.

Contents

List of Tables

List of Figures

Preface

An important problem of modern financial economics is understanding and quantifying the trade-off between risk and expected return. Although we anticipate the riskier assets to yield higher returns, the quantification of the risk-return trade-off was possible only after the introduction of the linear factor pricing models. Given the crucial role of these models in the asset pricing theory, my research analyzes the linear factor pricing models from both financial economics and econometrics points of view. On the financial economics side, I examine the role of time-varying market liquidity in explaining the future asset returns using a conditional multifactor asset pricing framework. On the econometrics side, I propose new distribution-free procedures for testing linear factor pricing models. Below I briefly elaborate on the three papers that constitute this dissertation.

First essay studies the role of time-varying market liquidity in explaining the time series behavior of the financial asset returns. Using lagged innovations in aggregate liquidity and its volatility as state variables, I investigate whether liquidity provides any information about the changes in assets' risk and returns over time, and whether this information can be attributed to a risk-based source. I start the empirical analysis by examining the effect of lagged innovations in liquidity and its volatility, at the portfolio level, without controlling for risk. Given that liquidity effect has a temporal dimension, I consider daily, weekly, and monthly data. The findings show that liquidity innovations track significant variation in portfolios' expected returns. The significant effect of liquidity shocks on portfolio returns can be attributed to one of the following sources: (1) A shock to aggregate liquidity first results in a change in the portfolio's risk, then the expected return. This risk-based view implies that liquidity *indirectly* affects the returns through time-variations in risk. (2) Liquidity shocks do not influence the risk, nevertheless, have a significant *direct* effect on the returns. In this case, liquidity provides information about future prices that is unrelated to risk, hence, it is a source of mispricing. To test whether liquidity variables remain to be significant after controlling for risk and determine the source of this effect, a conditional multifactor asset pricing model is adopted.

Employing the methodology of Shanken (1990), the conditional model allows both the expected returns and factor loadings to vary over time with liquidity innovations and volatility. The appeal of studying the conditional model in the time-series is two-fold. First, it allows us to examine the effect of aggregate liquidity variables on the time-varying portfolio risks. Second, since time-series analysis focuses on the expected returns, we can directly test whether the considered risk factors explain the variation in expected returns over time.

Second essay develops exact distribution-free tests of unconditional mean-variance efficiency. Empirical tests of the mean-variance efficiency hypothesis are usually conducted within the context of a multivariate linear regression. The application of tests based on asymptotic theory can lead to misleading conclusions as the approximation to the finite-sample distribution of test statistics can be quite poor, especially as the number of equations included in the system increases. The findings show that many standard parametric tests are unreliable, rejecting the null hypothesis of mean-variance efficiency far too often.

Without any parametric assumptions it would seem difficult to derive an exact finite-sample distribution theory. Despite this apparent difficulty, I propose in this paper new non-randomized tests that are exact in finite samples without any parametric assumptions about the distribution of the error terms in the simple multivariate linear regression model. Here I propose three testing approaches for joint inference on several parameters that differ mainly by what is assumed about the covariance of the errors across equations.

The first approach is an induced test procedure that allows for arbitrary covariances in the cross-section of error terms. The second approach assumes that the errors are independent across equations, conditional on the returns of the benchmark portfolio. The third approach is based on a simple linear combination of the test assets (or portfolios of test assets) and, like the second approach, provides a test procedure with the correct size no matter the number of included assets. These single-portfolio tests allow some forms of covariation in the cross-section of error terms. The number of assets in the cross-section may even exceed the number of time-series observations, making these tests particularly attractive when testing mean-variance efficiency with many test assets or when the portfolios have relatively short histories.

The proposed distribution-free (or non-parametric) tests have several appealing features, since they are built on the mere assumption that the joint temporal error density is symmetric around zero. This means that no restrictions are placed on the degree of non-normality or the degree of heterogeneity across marginal distributions. In fact, the existence of moments need not be assumed for the validity of the new tests. It is important to note that this framework still leaves open the possibility of asymmetries in the distribution of test asset returns via coskewness with the benchmark portfolio.

Asset returns typically display clear patterns of volatility clustering for which generalized autoregressive conditional heteroskedasticity (GARCH) models are often used. The tests proposed here allow not only for non-normalities, but also for *unknown* forms of conditional heteroskedasticity and other intertemporal dependencies among the absolute values of the error terms in the asset pricing model.

The third essay develops a finite-sample procedure to test the beta-pricing representation of linear factor pricing models that is applicable even if the number of test assets is greater than the length of the time series. Further, I make no parametric assumption about the distribution of the disturbances in the factor model.

This framework leaves open the possibility of unknown forms of time-varying non-normalities, heteroskedasticity, and even outliers in the asset returns.

I propose an adaptive approach based on a split-sample technique to obtain a single portfolio representation judiciously formed to avoid power losses that can occur in simple portfolio groupings. A very attractive feature of this approach is that it is applicable even if the number of test assets is greater than the length of the time series. This stands in sharp contrast to the standard test or any other approach based on usual estimates of the disturbance covariance matrix.

My proposed test procedure then exploits results from Coudin and Dufour (2009) to construct confidence sets for the model parameters by inverting exact sign-based statistics. The motivation for using this technique comes from an impossibility result due to Lehmann ans Stein (1949) that shows that the *only* tests which yield reliable inference under sufficiently general distributional assumptions, allowing non-normal, possibly heteroskedastic, independent observations are based on sign statistics.

The power of the proposed test procedure increases as either the times series lengthens or the cross-section becomes larger. Finally, I illustrate the new procedure by testing the well-known Fama-French factor model over 5-year subsamples of monthly returns on 100 U.S. equity portfolios formed on size and book-to-market.

Chapter 1

Time-Variation in Liquidity and Portfolio Returns

Abstract

This paper studies the role of time-varying liquidity in explaining the future asset returns. Using the innovations in liquidity and its volatility as liquidity risk measures, the methodology of Shanken (1990) is adopted in a multifactor asset pricing framework. The resulting conditional model allows the factor loadings to vary over time with the liquidity variables, hence, distinguish the risk and non-risk components of their explanatory power. The merit of the conditional model is its ability to directly test whether the innovations in liquidity and volatility capture time-variation in the risk of an asset and whether they contain further information after controlling for changes in risk. Using daily, weekly, and monthly data for the period of January 1964 - December 2008, I find that both the innovations in liquidity and its volatility are strongly associated with changes in assets' risk. After controlling for the time-varying risk, the liquidity variables have no impact on the expected returns at the weekly and monthly horizons. However, the daily innovations in liquidity convey information about the future prices beyond the risk explanation.

JEL classification: G12

Keywords: Liquidity, Factor Models

1.1 Introduction

The liquidity of an asset is often defined as the ability to cheaply trade large quantities in a short period of time without moving the price too much. Despite its measurement difficulty, there is substantial evidence to show that both the level of liquidity and liquidity risk are priced in the market. Focusing on the latter, this paper examines the role of time-varying liquidity in explaining the future asset returns. More specifically, using lagged innovations in aggregate liquidity and its volatility as state variables, I investigate whether liquidity provides any information about the changes in assets' risk and returns over time, and whether this information can be attributed to a risk-based source.

The early studies, such as Amihud and Mendelson, (1986), Brennan and Subrahmanyam (1996), Brennan, Chordia, and Subrahmanyam (1998), Datar, Naik, and Radcliffe (1998), examine the cross-sectional relation between the level of liquidity and expected returns. They find that expected returns are decreasing in liquidity. More recent studies by Chordia, Roll, and Subrahmanyam (2000), Hasbrouck and Seppi (2001), Huberman and Halka (2001) analyze the systematic component in liquidity and show that the time-variation in liquidity exhibits commonality. The new evidence of a systematic component suggests that the fluctuations in liquidity may be a priced risk factor. In general, the trading costs are high when aggregate liquidity is lower. These higher costs are especially unwelcome to an investor whose wealth has already dropped and has a higher marginal utility of wealth. Therefore, investors must require additional compensation for being exposed to liquidity risk due to its variations over time. Motivated by the risk view of liquidity, Pastor and Stambaugh (2003), Acharya and Pedersen (2005), Liu (2006), and Watanabe and Watanabe (2008) investigate market-wide liquidity as a state variable that is common to all stocks, rather than a characteristic that is relevant for pricing. The cross-sectional results of these studies provide evidence that liquidity is a source of priced systematic risk in stock returns. A related study by Chordia, Subrahmanyam, and Anshuman (2001) examines the effect of the volatility of liquidity on stock returns, finding a surprising negative association in the cross-section.

Although the cross-sectional relation between liquidity and expected returns has been widely investigated, there is only limited time-series evidence on this issue. Moreover, the existing literature do not seem to agree on the direction of the time-series relation. Using monthly and yearly data Amihud (2002) finds that liquidity and future portfolio excess returns are negatively related. Jones (2002) also confirms the same negative relation using a very long time-series. However, the short-term results of Gervais, Kaniel, and Mingelgrin (2002) show that the relation between liquidity and expected returns is positive up to 20 days. The same positive relation is then verified for international developed markets by Kaniel, Li, and Starks (2005). These mixed time-series results indicate that, the relation between liquidity and expected returns has a temporal dimension.

The inconclusive time-series evidence and the growing interest in market-wide liquidity risk motivated me to address the research question: What is the role of time-varying aggregate liquidity on assets' overall risk, measured by factor loadings in a conditional asset pricing model, and expected returns? Employing the illiquidity ratio of Amihud (2002), the liquidity risk is proxied by the innovations in aggregate liquidity and its time-varying volatility. In a similar spirit, Watanabe and Watanabe (2008) examined the time-variation in liquidity risk in a bivariate regime-switching model. In this paper, however, instead of investigating the effect of liquidity only on liquidity risk, I study its effect on three Fama and French (1993) factors and a liquidity factor.

I start the empirical analysis by examining the effect of lagged innovations in liquidity and its volatility, at the portfolio level, without controlling for risk. Given that liquidity effect has a temporal dimension, I consider daily, weekly, and monthly data. The findings show that liquidity innovations track significant variation in portfolios' expected returns. The significant effect of liquidity shocks on portfolio returns can be attributed to one of the following sources: (1) A shock to aggregate liquidity first results in a change in the portfolio's risk, then the expected return. This risk-based view implies that liquidity *indirectly* affects the returns through time-variations in risk. (2) Liquidity shocks do not influence the risk, nevertheless, have a significant *direct* effect on the returns. In this case, liquidity provides information about future prices that is unrelated to risk, hence, it is a source of mispricing. To test whether liquidity variables remain to be significant after controlling for risk and determine the source of this effect, a conditional multifactor asset pricing model is adopted. The controlled risk factors in this model include the three Fama and French (1993) factors and a liquidity factor, as suggested in Pastor and Stambaugh (2003), Liu (2006), and Miralles and Miralles (2006). Similar to Fama and French's SMB (small minus big) and HML (high minus low), the mimicking liquidity factor is constructed based on the return differences between an illiquid stocks portfolio and a liquid stocks portfolio.

Employing the methodology of Shanken (1990), the conditional model allows both the expected returns and factor loadings to vary over time with liquidity innovations and volatility. In this regression equation, the variation in the intercept measures the direct effect of liquidity variables that is unrelated to risk. On the other hand, if the liquidity innovations and volatility are truly risk factors, their effect will be captured by the varying factor loadings and the resulting intercept will be zero. The appeal of studying the conditional model in the time-series is two-fold. First, it allows us to examine the effect of aggregate liquidity variables on the time-varying portfolio risks. Second, since time-series analysis focuses on the expected returns, we can directly test whether the considered risk factors explain the variation in expected returns over time.

This paper is organized as follows: Section 2 provides the theoretical background about the relation between liquidity and asset prices. Section 3 introduces the data and time-series methodology. The employed models, and the details about obtaining the innovations in liquidity and time-varying volatility are explained in this section. Section 4 presents the empirical results for a system of predictive regressions, an unconditional model, and a conditional four-factor asset pricing model. Section 5 concludes.

1.2 Theoretical Background

The standard asset pricing theory lies on two crucial assumptions; *frictionless markets* and *no arbitrage*. In our context, the existence of frictionless markets implies that trading assets in the stock market does not involve transaction costs, in other words, stocks are perfectly liquid at all times. As a result, the standard theory shows that stocks with the same cash flows must have the same price, otherwise there will be arbitrage opportunities available to the investors at no risk. For an investor who can freely trade in these frictionless markets, the first-order condition describing the consumption and portfolio plan is $P_t = E_t[P_{t+1}M_{t+1}]$. The stochastic discount factor or the pricing kernel, M_{t+1}, summarizes all the necessary information for the asset price P_t.

The assumption of frictionless markets, however, is rather extreme. For instance, while trading in the stock market, the investors face transaction costs that limit their ability to exploit return patterns. Hence, different prices for assets with the same cash flows do not necessarily imply the existence of arbitrage (see Amihud and Mendelson, 1986; Silber, 1991; Brenner et.al., 2001). Rather it implies that the stochastic discount factor that prices all the assets actually does not exist. Some liquidity models, however, continue to rely on the existence of the stochatic discount factor but assume that it is a function of the aggregate liquidity, i.e. Pastor and Stambaugh (2003).

Relaxing the assumption of frictionless markets, the theory of liquidity-based asset pricing suggests that the level of liquidity as well as the liquidity risk are priced in the market. The intuition behind the effect of liquidity is that the price of a stock is the discounted value of the entire future stream of cash flows net of the transaction costs, such as:

$$P_i = \frac{\bar{C}_i - \mu A_i}{r_f} \tag{1.2.1}$$

where P_i is the stationary equilibrium price, $\bar{C}_i$ is the mean of the i.i.d. cash-flows, and A_i is the cost of transaction of asset i. The trading intensity is denoted by μ, showing how often the investor expects to trade asset i. Finally r_f is the risk-free rate.[1] Similarly, the required return on stock i can be written as:

$$E(r_i) = r_f + \mu \frac{A_i}{P_i} \tag{1.2.2}$$

Eq. (1.2.1) and (1.2.2) show how the level of liquidity affects the stock prices and required returns. Intuitively, Eq. (1.2.2) simply demonstrates that the required return on stock i is the sum of risk-free rate and the expected per period cost of trading. Moreover, the liquidity risk, which arises due to the variation of liquidity over-time, is expected to be priced in the market. The variation in liquidity is perceived as risk because its fluctuations are correlated with the price volatility and also it creates additional uncertainty about the future transaction costs that the investor will face.

Separate from the effect of liquidity on asset returns net of transaction costs, a fairly new line of research has emerged suggesting that liquidity is a common risk factor (see; Pastor and Stambaugh, 2003; Acharya and Pedersen, 2005; Liu, 2006; Miralles and Miralles, 2006). Based on the observation that the time-variation in liquidity exhibits commonality (see; Chordia, Roll, and Subrahmanyam, 2000, Hasbrouck and Seppi, 2001, Huberman and Halka, 2001), these studies show that investors must be compensated for holding stocks with high sensitivity to innovations

[1] The future stream of cash flows and transaction costs is divided by the risk-free rate, r_f, because the discount rate for a risk-neutral investor is $1/R_f$, with $R_f = 1 + r_f$.

in aggregate liquidity.

1.3 Data and Methodology

The empirical results are based on daily data for stock return, price, and trading volume from daily stock files of Chicago University's Center for Research of Securities Prices (CRSP). The sample consists of all the ordinary common stocks (CRSP share codes 10 and 11) listed in New York Stock Exchange (NYSE) and American Stock Exchange (AMEX) over the period January 1964 to December 2008. To be admitted in the sample, a stock must have a beginning of the month price between $5 and $1000. Moreover, each stock is required to have at least 15 days of return and volume data in each month in a year and at least 100 days of data in the previous year. Following Amihud (2002), the illiquidity measure for a stock is defined as the ratio of the stock's absolute return to the dollar value of its trading volume (multiplied by 10^6):

$$ILLIQ_{id} = \frac{|R_{id}|}{VOL_{id}} * 10^6 \tag{1.3.1}$$

where R_{id} is the return and VOL_{id} is the dollar volume of stock i on day d. The illiquidity measure in Eq. (1.3.1) focuses on the stock price fluctuations due to order flow. Intuitively, it measures the effect of trading volume on the stock price. The larger the response of the stock price to the trading volume (a high value of $ILLIQ_{id}$)

the less liquid the stock i is, therefore, $ILLIQ$ is referred as *illiquidity ratio*. Then, the stocks are sorted into 10 illiquidity portfolios each year based on their average illiquidity ratio at the end of the previous year. The stocks with highest $ILLIQ_{id}$ are allocated to the least liquid portfolio P_1, and the stocks with lowest $ILLIQ_{id}$ are allocated to the most liquid portfolio P_{10}. The illiquidity for a portfolio is calculated as the natural logarithm (ln) of equally-weighted average of the illiquidity ratios for individual stocks in that portfolio. Using daily return and volume data, the daily cost of trading for any portfolio p with I stocks is:

$$ILLIQ_{pd} = ln\left(\frac{1}{I}\sum_{i=1}^{I}\frac{|R_{id}|}{VOL_{id}}\right) \tag{1.3.2}$$

The market portfolio, denoted by P_m, is defined as the portfolio of all available stocks, hence, the market-wide illiquidity is the average of the illiquidity ratios for all stocks. Based on the previous research, Amihud's illiquidity ratio have theoretical and empirical advantages over the usual measures. Hasbrouck (2004) finds that, among the liquidity proxies he considered, Amihud's appears to be the best that captures Kyle's lambda. Amihud (2002) shows that, $ILLIQ$ is positively related with measures of price impact which he obtained from microstructure data. Similarly, Sadka (2006) finds that, illiquidity ratio is highly correlated with high-frequency measures of price impact. It also has empirical advantage, because the required data is relatively easy to obtain. The availability of data for a long period of time allows us to examine the time-series variations in liquidity and its effects on expected returns. The empirical results in this paper are based on the assumption that, $ILLIQ$ is a valid proxy for illiquidity costs.

To investigate the relation between liquidity, expected returns, and common risk factors in an asset pricing framework, I obtained the three Fama-French factors; market (MKT), size, (SMB), and book-to-market (HML), as well as the risk-free one-month Treasury-bill rate from Kenneth French's website.[2] I also created a liquidity-mimicking factor based on $ILLIQ$. The liquidity factor is constructed by sorting all the available stocks into three illiquidity groups based on NYSE breakpoints for the top 30% (illiquid), middle 40% (moderately liquid), and bottom 30% (liquid). Then, the liquidity factor IML (illiquid minus liquid) is defined as the difference between the mean returns of the illiquid stocks portfolio and liquid stocks portfolio. The presented portfolio returns are equally-weighted returns.

To study the possible temporary relation between the interest variables, I aggregated the daily data to weekly and monthly frequencies. The lower frequency data for portfolio returns and IML liquidity factor are obtained by aggregating the daily observations. The illiquidity ratio is aggregated by averaging the daily observations

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

across dates.

Table 1.1 reports the summary statistics for four risk factors, MKT, SMB, HML, and IML at each data frequency. For simplicity, I will focus on the monthly data in my explanations, unless it is necessary to point out the temporary aspect in results. The average monthly returns on MKT, SMB, HML, and IML are 0.37%, 0.26%, 0.43%, and 0.44% respectively. The positive risk premiums for each factor show that investors are compensated for bearing factor risk. Also, notice that IML provides the highest premium implying the importance of the liquidity risk factor.

A single factor model, such as Capital Asset Pricing Model (CAPM), studies the relationship between an asset's market beta and expected return. To ensure that the market beta from a multifactor asset pricing model has the same interpretation as the one from a single factor model, the included risk factors must be orthogonal to each other. The correlation coefficients among MKT, SMB, HML, and IML are reported in Table 1.1. The non-zero values indicate that the data on each factor fails to control for the effects of the other factors. At monthly frequency, Panel C, SMB and IML are positively correlated with MKT (correlation coefficients equal to 0.30 and 0.12 respectively), while HML has a negative correlation of -0.38. The correlation between SMB and HML is -0.26, and the highest correlation is between SMB and IML with a positive value of 0.84. The four risk factors are orthogonalized using the following regressions:

$$SMB_t = A^{smb} + B^{smb} MKT_t + \epsilon_t^{smb} \quad (1.3.3)$$

$$HML_t = A^{hml} + B^{hml} MKT_t + C^{hml} SMB_t^{\perp} + \epsilon_t^{hml} \quad (1.3.4)$$

$$IML_t = A^{iml} + B^{iml} MKT_t + C^{iml} SMB_t^{\perp} + D^{iml} HML_t^{\perp} + \epsilon_t^{iml} \quad (1.3.5)$$

where the orthogonalized factors, denoted by superscript $\perp$, are the sum of the estimated residuals and the intercept from Eq. (5)-(7), such that:

$$SMB_t^{\perp} = \hat{A}^{smb} + \hat{\epsilon}_t^{smb} \quad (1.3.6)$$

$$HML_t^{\perp} = \hat{A}^{hml} + \hat{\epsilon}_t^{hml} \quad (1.3.7)$$

$$IML_t^{\perp} = \hat{A}^{iml} + \hat{\epsilon}_t^{iml} \quad (1.3.8)$$

Table 1.1 shows that the cross-correlations across the orthogonalized risk factors are all equal to zero. Hence, in the calculation of a risk factor, the above procedures successfully control for the impact of the other factors. The monthly mean return on the size and the liquidity factors decline from 0.26% to 0.18% and from 0.44% to 0.12% respectively. The monthly average return on the book-to-market factor, however, increases from 0.43% to 0.54%.

1.3.1 Aggregate Illiquidity Ratio

Fig. 1.1 plots the time-series of illiquidity ratios for the market portfolio (P_m), an illiquid portfolio (P_2), and a liquid portfolio (P_9) at daily, weekly, and monthly frequencies. The most prominent feature of the illiquidity series is the downward trend, indicating that liquidity has increased over the 45-year sample period. The liquidity of P_9 has increased faster, resulting in a larger liquidity differential between the liquid and illiquid portfolios. Moreover, the decidedly negative trend of illiquidity ratio reveals that, the series is non-stationary.

If the measure of aggregate illiquidity is a good proxy, then it is expected to capture the liquidity conditions in the market and coincide with the major economic and financial events. Fig. 1.2 plots the monthly aggregate illiquidity and the equally-weighted returns of all the available NYSE-AMEX stocks. The gray areas show the NBER business-cycle contraction dates, and each black line represents a large decline in liquidity due to a major event. The illiquidity series presents that, the economic slow-downs are accompanied by large tightening in aggregate liquidity, except the illiquidity spike leads the slow-down between March 2001 and November 2001. Throughout the sample period from January 1964 to December 2008, the large liquidity declines coincide with six major events including the Penn Central Debacle in May 1970, the oil crisis in November 1973, stock market crash in October 1987, the decline in stock market in November 1990 due to the turmoil in Persian Gulf, the liquidity decline in September 1998 following Asian crisis, Russian default, and the collapse of the Long Term Capital Management, and finally, the plunge in the stock market in October 2008 as a result of failures in large U.S. financial institutions because of subprime loans and credit default swaps.

Table 1.2 reports the summary statistics of average returns and illiquidity ratios for ten liquidity-based portfolios, P_1-P_{10}, and the market portfolio, P_m, at each data frequency. The statistics in Panel A through Panel C show that, portfolio returns and illiquidity ratios at each frequency follow similar patterns. For instance, the unconditional average monthly return for individual portfolios is decreasing in liquidity. The average monthly return of P_1 is 1.817% and of P_{10} is 0.854%, for an annualized spread of 12.19%.[3] This economically significant return spread among the liquidity portfolios confirms the existence of liquidity premium.

Table 1.2 also more formally documents the previously observed nonstationarity in illiquidity series. The estimated first-order autocorrelation coefficients reveal that, the series is highly persistent (see also; Amihud, 2002; Eckbo and Norli, 2002; Watanabe and Watanabe, 2008), with values equal to 0.98, 0.99, and 0.99 for the market portfolio P_m at daily, weekly, and monthly frequencies respectively. To remove this persistence in the illiquidity and obtain a time-series of liquidity innovations, in the next section

[3] I annualized the daily, weekly, and monthly returns using $[(1 + R/100)^k - 1] * 100$, where k is the number of observations in one year.

I take the first difference of the aggregate illiquidity ratio.

1.3.2 Aggregate Liquidity Innovations and Volatility

The change in the aggregate illiquidity at the end of each period is defined as the first-difference of the market-wide illiquidity ratio, $\Delta ILLIQ_{mt} = ILLIQ_{mt} - ILLIQ_{m,(t-1)}$, for $t = 1, ..., T$. In a rational expectations model, the expected changes in state variable(s) should not affect the risk premium. In other words, only the unexpected changes in liquidity result in an adjustment about the future liquidity expectations, hence the risk premium. Consistent with the theory, many studies in the literature employ the innovations in liquidity as a measure of liquidity risk (see Pastor and Stambaugh, 2003; Liu, 2006; Chan et. al., 2008; and Watanabe and Watanabe, 2008)[4].

The relation between excess stock returns and liquidity innovations would be simply analyzed in a linear regression framework, if the latter was observable. Since this is not the case, the common method used in the literature is a two-step estimation ($2SE$) procedure. In the first-step, the changes in the illiquidity measure are decomposed into expected and unexpected components using a simple $AR(1)$ model;

[4]The empirical results using the data in this paper also show that, the expected component of illiquidity ratio has no significant effect on the future portfolio returns. The unreported results are available at request.

$\Delta ILLIQ_{mt} = \alpha + \rho\, \Delta ILLIQ_{m,(t-1)} + \eta_{mt}$, where the estimated residuals ,$\hat{\eta}_{mt}$, represent the unexpected changes in illiquidity. In the second-step, the excess stock returns are regressed on the estimated residuals from the first-step, replacing the unobserved illiquidity innovations. The generated regressors in this fashion, however, cause two issues: (i) Even if the estimates from the first-step are consistent, the presence of the generated regressors at the second step generally leads to loss of efficiency in estimation. Moreover, it results in invalid inference, because the obtained standard errors are an inconsistent estimate of the true standard errors (see Pagan, 1984; 1986) [5]. (ii) The second issue concerns the information used to estimate the residuals from an $AR(1)$ model. The parameter estimates from the above $AR(1)$ model are obtained by using all the available information up to end of the data sample (period T). Therefore, the estimated residuals at time t, $\hat{\eta}_{mt} = \Delta ILLIQ_{mt} - \hat{\alpha} - \hat{\rho}\, \Delta ILLIQ_{m,(t-1)}$, do not actually represent the residuals for that time, since, $\hat{\alpha}$ and $\hat{\rho}$ are obtained using information available before, as well as, after time t. To address this problem,

[5] To avoid the inefficiency in the parameter estimates and inconsistency in the standard errors, I have estimated the models in section (3.3) in one-step. In this system of nonlinear least squares (NLS) the parameters of the first- and second-step from $2SE$ are simultaneously estimated. However, this procedure requires simultaneous estimation of 42 parameters for the simple predictive regressions in Eq. (1.3.14), and 152 parameters for the conditional multifactor asset pricing model in Eq. (1.3.15). Due to the large number of parameters, the variance-covariance matrices from the NLS procedure became intractable.

I estimate a recursive $AR(1)$ model for changes in the aggregate illiquidity:

$$\Delta ILLIQ_{mt} = \alpha_t + \rho_t \Delta ILLIQ_{m,(t-1)} + \eta_{mt} \ , \quad t = 1, ..., T \tag{1.3.9}$$

where the intercept and the slope coefficient on $\Delta ILLIQ_{m,(t-1)}$ are not constant but evolve over time. The recursive $AR(1)$ process in Eq. (1.3.9) implies that, the agents learn about the illiquidity process over time. They believe that, market illiquidity follows the process in Eq. (1.3.9) but they do not know the actual values for α_t and ρ_t. Each period they update their parameter estimates incorporating the new information in that period.

The recursive estimation procedure starts estimating the parameters of Eq. (1.3.9) by using the first 30 observations of changes in illiquidity. Then, the procedure is repeated for the next period, by including the data for the new date. It continues until the end of the dataset is reached. At the end of this process two series, one for $\hat{\alpha}_t$ and one for $\hat{\rho}_t$, are generated with $\tilde{T} = (T - 30)$ estimates each. The innovations in market liquidity, denoted by $\hat{\eta}^*_{mt}$, is defined as the negative values of the residuals from Eq. (1.3.9)[6]:

[6] The residuals from Eq. (1.3.9) are multiplied by negative one to enhance the interpretation of results by converting them into a measure of liquidity shocks rather than illiquidity.

$$\begin{aligned} \hat{\eta}^*_{mt} &= -(\Delta ILLIQ_{mt} - \hat{\alpha}_t - \hat{\rho}_t \Delta ILLIQ_{m,(t-1)}), \quad t = 1, ..., \tilde{T} & (1.3.10) \\ &= -\hat{\eta}_{mt} & (1.3.11) \end{aligned}$$

Note that, a higher positive value of $\hat{\eta}^*_{mt}$ now represents innovations in aggregate liquidity. The time-series plot of liquidity innovations in Fig. 1.3 appears to have constant mean at each data frequency.

In addition to the innovations in market-wide liquidity, this paper also examines the *volatility* of liquidity as a state variable for predicting excess returns. The plots in Fig. 1.3 exhibits time-varying volatility clustering in liquidity innovations. Hence, the estimated residuals from Eq. (1.3.9) are defined as $\hat{\eta}_{mt} = \sigma_{mt} u_{mt}$, where $u_{mt} \sim D(0,1)$, and D specifies a distribution with mean zero and unit variance. Based on the Akaike information criterion (AIC), the time-varying volatility of daily liquidity is modeled as a $GARCH(1,1)$ process:

$$\sigma^2_{mt} = \delta_{0t} + \delta_{1t} \hat{\eta}^2_{m,(t-1)} + \delta_{2t} \sigma^2_{m,(t-1)} \tag{1.3.12}$$

and the volatility of weekly, and monthly liquidity are modeled as $ARCH(2)$ processes:

$$\sigma_{mt}^2 = \delta_{0t} + \delta_{1t}\hat{\eta}_{m,(t-1)}^2 + \delta_{2t}\hat{\eta}_{m,(t-2)}^2 \tag{1.3.13}$$

Notice that the parameters in Eq. (1.3.12) and (1.3.13) have time subscripts. It is because the square-root of the estimated variances from the above procedures, $\hat{\sigma}_{mt}$, will be employed in the next section as predictors of the excess portfolio returns. However, since these series are estimated, one more time we face the previously discussed generated repressors problem. This issue is addressed by recursively estimating the above $GARCH$ and $ARCH$ processes. First, using 1000 observations for $GARCH(1,1)$ of daily volatility and 50 observations for $ARCH(2)$ of weekly and monthly volatilities, the initial $\hat{\delta}_t$ estimates are obtained. Then, the process is repeated by including the $\hat{\eta}_{mt}$ for the next period, until the end of the dataset is reached. The resulting number of estimates for each $\hat{\delta}_t$ series is equal to $T^* = (\tilde{T} - B)$, where B is the number of observations employed for the initial estimates.

1.3.3 Time-Series Methodology

The methodology introduced in this section initially examines the simple time-series relation between the two liquidity variables, namely innovations in aggregate liquidity and its volatility, and future portfolio returns. The following system of regression equations is employed for ten liquidity-sorted portfolios:

$$Z_{pt} = \beta_{p0} + \beta_{p1}\hat{\eta}^{*}_{m,(t-1)} + \beta_{p2}\hat{\sigma}_{m,(t-1)} + \beta_{p3}Jan_t + \varepsilon_{pt}, \qquad p = 1, ..., 10 \qquad (1.3.14)$$

where $Z_{pt} = R_{pt} - R_{ft}$ is the portfolio p's return in excess of risk-free rate R_f and Jan_t is a dummy variable for the month of January. The variable Jan controls for the January effect in liquidity portfolios, which originally refers to the fact that, the returns of small firms are higher in the month of January than in other months (see Keim; 1983). Eq (1.3.14) provides a system of equations, where a separate time-series regression is defined for each portfolio. It only serves the purpose of understanding the basic relation between excess returns and lagged liquidity variables. It does not control for risk factors and does not supply any information about the source of time-varying expected returns.

The efficient market hypothesis states that the liquidity variables should not have any predictive power after controlling for all the risk factors. In other words, the considered factors are adequate proxies for priced risk, and liquidity shocks and volatility capture information about changes in these factors. Otherwise, aggregate liquidity variables will remain to be significant even after controlling for risk. The latter outcome implies that liquidity variables do not contribute to the portfolios' risk. However, they capture non-risk related information (mispricing) about expected returns.

I examine the source of time-varying expected returns by using the conditional methodology developed by Shanken (1990) for a four-factor asset pricing model. The conditional model controls for the orthogonalized versions of the three Fama and French (1993) factors, the excess market return (MKT), size factor ($SMB^{\perp}$), and book-to-market factor ($HML^{\perp}$). In addition, recognizing the limitations of three-factor model[7], I include a liquidity factor $IML^{\perp}$ (illiquid minus liquid) as in Liu (2006) and Miralles and Miralles (2006). The four-factor model assigns a role for liquidity as a common risk factor similar to $SMB^{\perp}$ and $HML^{\perp}$. The liquidity factor $IML^{\perp}$ is created by taking the return differences between the lowest liquidity and the highest liquidity portfolios. The resulting conditional four-factor model takes the following form:

$$Z_{pt} = a_{pt} + b_{pt}MKT_t + s_{pt}SMB_t^{\perp} + h_{pt}HML_t^{\perp} + d_{pt}IML_t^{\perp} + e_{pt} \tag{1.3.15}$$

The factor loadings in the system of equations (1.3.15) vary over time with the innovations in liquidity and volatility. Following Shanken (1990), I assume that they

[7]Daniel and Titman (1997) argue that the Fama and French factors do not have the ability to directly explain the cross-section of average returns. They claim that the explanatory power of the Fama-French factors is not due to their ability to capture risk but due to their correlation with the firms' characteristics. Another drawback of the three-factor model for our specific purpose is its lack of control for the liquidity risk.

are linearly related, such that:

$$
\begin{aligned}
a_{pt} &= a_{p0} + a_{p1}\hat{\eta}^*_{m,(t-1)} + a_{p2}\hat{\sigma}_{m,(t-1)} && (1.3.16)\\
\lambda_{pt} &= \lambda_{p0} + \lambda_{p1}\hat{\eta}^*_{m,(t-1)} + \lambda_{p2}\hat{\sigma}_{m,(t-1)} && (1.3.17)
\end{aligned}
$$

where a_{pt} is the intercept and $\lambda_{pt} = (b_{pt}, s_{pt}, h_{pt}, d_{pt})'$ is the vector of factor loadings in Eq (1.3.15). Then Eq. (1.3.15) can be rewritten in a more explicit form:

$$
\begin{aligned}
Z_{pt} = \; & a_{p0} + a_{p1}\hat{\eta}^*_{m,(t-1)} + a_{p2}\hat{\sigma}_{m,(t-1)} \\
& + \left(b_{p0} + b_{p1}\hat{\eta}^*_{m,(t-1)} + b_{p2}\hat{\sigma}_{m,(t-1)}\right) MKT_t \\
& + \left(s_{p0} + s_{p1}\hat{\eta}^*_{m,(t-1)} + s_{p2}\hat{\sigma}_{m,(t-1)}\right) SML_t^{\perp} \qquad (1.3.18)\\
& + \left(h_{p0} + h_{p1}\hat{\eta}^*_{m,(t-1)} + h_{p2}\hat{\sigma}_{m,(t-1)}\right) HML_t^{\perp} \\
& + \left(d_{p0} + d_{p1}\hat{\eta}^*_{m,(t-1)} + d_{p2}\hat{\sigma}_{m,(t-1)}\right) IML_t^{\perp} + e_{pt}, \qquad p = 1, ..., 10
\end{aligned}
$$

Eq. (1.3.18) contains fifteen independent variables; an intercept, four factors, and ten interactive terms. The coefficients a_{p1} and a_{p2} measure the *direct effect* of lagged liquidity innovations and volatility respectively. The changes in the factor loadings due to liquidity variables are captured by $\lambda_{p1} = (b_{p1}, s_{p1}, h_{p1}, d_{p1})'$ and $\lambda_{p2} = (b_{p2}, s_{p2}, h_{p2}, d_{p2})'$. Using the conditional multifactor model, I can directly test the source of the time-variation in the expected portfolio returns by distinguishing the risk and non-risk components. The rational asset pricing theory states that, the effect of liquidity should disappear once all the risks are controlled for. Within the context of Eq. (1.3.18), there are mainly two possible sources of the varying expected returns. First, if the non-risk related component, measured by a_{p1} and a_{p2}, is significantly different from zero then the *direct effect* of liquidity prevails even after controlling for the the time-variation in the portfolios' risk. Assuming that the considered factors adequately control for risk, such an outcome will imply that, the risk factors do not fully explain the time-series relation between $\hat{\eta}_m^*$, $\hat{\sigma}_m$, and expected returns, hence, the rational asset pricing hypothesis does not hold. Second, if the estimated values of a_{p1} and a_{p2} are zero then the liquidity variables have an effect through the time-variation in factor loadings. In this case, the source of the varying expected returns will be attributed to changes in the portfolio's risk, verifying the rational asset pricing theory. The next section provides the test results for the relation between liquidity variables and expected returns, and the source of this relation.

1.4 Empirical Results

To gain insight about the time-series relation between the innovations in liquidity ($\hat{\eta}_m^*$), its volatility ($\hat{\sigma}_m$), and the expected returns, I start my empirical analysis by estimating the seemingly unrelated regressions in Eq. (1.3.14). This system of equations for ten liquidity-sorted portfolios evaluates the economic importance of $\hat{\eta}^*_m$ and $\hat{\sigma}_m$ without controlling for risk. Because each equation in the system has the same regressors, the least-squares method provides an efficient estimator, hence, it is favored. Table 1.3 reports the results for each portfolio. The averages of the coefficient estimates across portfolios and the Wald statistics are reported in the last two columns. The Wald statistic tests the null hypothesis that, the estimates of a coefficient for all portfolios are jointly equal to zero. It is calculated as $W_k = \hat{\beta}'_{pk}(\hat{\Sigma}_k)^{-1}\hat{\beta}_{pk}$, where $\hat{\beta}_{pk} = (\hat{\beta}_{1k}, ..., \hat{\beta}_{10,k})'$ is the vector of β_{pk} estimates for the k^{th} regressor from ten equations, and $\hat{\Sigma}_k$ is the (10×10) heteroskedasticity-consistent estimate of the residual covariance matrix. Under the null hypothesis, the Wald statistic is asymptotically distributed as χ^2 with degrees of freedom equal to the number of restrictions.

The evidence in Panel A through C shows that, the lagged innovations in aggregate liquidity has a significant positive impact on the expected portfolio returns at all data frequencies. At first, a positive relation between lagged liquidity and expected returns seems surprising. However, the positive sign in due to the estimated negative $\hat{\rho}_t$ values from Eq. (1.3.9). The autocorrelation coefficient in Eq. (1.3.9) is time-varying with a mean of -0.06. Together with the persistence in liquidity (see section 3.1) the negative $\hat{\rho}_t$ estimates imply that a negative (positive) liquidity shock in period t is expected to be followed by a low (high) liquidity level (due to persistence) in period $t+1$, but is still expected to be higher than the level in period t (due to negative $\hat{\rho}_t$). Therefore, an unexpected decline (increase) in liquidity will result in a lower expected return in the next period. On the other hand, the volatility of liquidity appears to be insignificant for all portfolios at all frequencies.

For daily data (Panel A), the coefficients of the lagged liquidity innovations range from -0.001 for P_{10} to 0.168 for P_1. Although, the effect of $\hat{\eta}^*_m$ is stronger on less liquid portfolios, it remains to be significant for P_1 through P_6. The coefficients of $\hat{\eta}^*_m$ have a positive average of 0.099 across portfolios. Hence, on average, an unexpected increase in aggregate liquidity at time t results in higher expected returns at time $t+1$. Together with the standard deviation of 0.048, the average coefficient estimate of $\hat{\eta}^*_m$ shows the magnitude of its impact. For example, an increase in $\hat{\eta}^*_m$ two standard deviations from its mean implies a 0.01% (0.096*0.099) daily, 2.42% annualized, increase in the next period return of a typical portfolio. Supporting evidence for the significant effect of $\hat{\eta}^*_m$ is also provided by the Wald test statistic, W. At 1% significance level, it strongly rejects the null hypothesis with a value of 35.68 and confirms that the aggregate liquidity shocks co-move with the expected returns. The weekly results in Panel B and the monthly results in Panel C are similar to the daily ones. $\hat{\eta}^*_m$ is positively related with the expected returns of the liquidity-based portfolios, and the degree of association is higher for less liquid portfolios. The average coefficient estimate is 0.525 for weekly, and 3.304 for monthly frequencies, with respective standard deviations of 0.550 and 1.195. The implied effect of a two standard deviations increase in $\hat{\eta}^*_m$ on the next period return is 0.58% at weekly and 7.90% at monthly horizons. The Wald test statistic again rejects the null hypothesis at 1% significance level.

The least-squares estimation method assumes that the regressors $\hat{\eta}^*_m$ and $\hat{\sigma}_m$ are exogenous. This assumption may not hold if there are feedback effects from the expected portfolio returns to liquidity variables. Since Eq. (1.3.14) specifies a system of predictive regressions with *lagged* innovations and volatility, endogeneity in the form of feedback is less of a problem. However, if the investors' liquidity preferences are affected by the expectations of future prices, then the feedback effects can still exist. To investigate the latter, I estimate the correlations between the contemporaneous liquidity innovations and the residuals from Eq. (1.3.14). The low values, ranging between 0.10 - 0.13 for daily, 0.23 - 0.30 for weekly, and 0.42 - 0.54 for monthly data, suggest that the significant effect of $\hat{\eta}^*_m$ on expected returns is not due to endogeneity bias.

The lagged innovations in liquidity significantly affects the expected returns, however, provides low R^2 statistics. Although, the small values of R^2 are common in predictive regressions of expected returns, they still cast doubt about the economic importance of $\hat{\eta}^*_m$. Campbell and Thompson (2008) explore this issue in a simple optimal portfolio problem and conclude that, low R^2 statistics can generate large benefits for investors. They compare the average excess returns from a conditional portfolio of a mean-variance optimizing investor who observes the predictor variable(s) and an unconditional portfolio of an investor who does not. They show that the proportional increase in expected return from observing the predictor(s) is approximately equal to R^2/S^2, where S is the unconditional Sharpe ratio of the risky asset. Table 1.4 reports the unconditional Sharpe ratios, R^2 statistics from Eq. (1.3.14), and the R^2/S^2 ratios for ten liquidity-sorted portfolios. The higher values of R^2/S^2 ratio for illiquid portfolios imply that the liquidity variables provide more information for these portfolios. Nevertheless, the R^2/S^2 ratios are high for the majority of the portfolios. For example, the daily squared Sharpe ratio for P_8 is 0.0005 (0.05%), and the R^2 equals to 0.001 (0.1%). The resulting $R^2/S^2 = 2.0\%$ means that using the information from liquidity variables, $\hat{\eta}_m^*$ and $\hat{\sigma}_m$, a mean-variance optimizing investor can obtain a 2.0% proportional increase in her daily return. The weekly and monthly proportional return gains for the same portfolio are 0.8% and 1.5% respectively.

1.4.1 The Unconditional Four-Factor Model

The next section examines the source of time-variation in expected portfolio returns in a four-factor asset pricing model. In addition to the orthogonalized three Fama and French (1993) factors, this model includes a liquidity factor ($IML^{\perp}$). To determine whether $IML^{\perp}$ provides relevant information for portfolio returns, I investigate its significance in an unconditional framework.

Table 1.5 reports the least-squares estimates for the unconditional system of equations; $Z_{pt} = a_p + b_p MKT_t + s_p SMB_t^{\perp} + b_p HML_t^{\perp} + d_p IML_t^{\perp} + e_{pt}$, where $p = 1, ..., 10$. Panels A through C document the results for daily, weekly, and monthly data. Consistent with the previous findings in the literature, the factor loadings on size ($SMB^{\perp}$) and book-to-market ($HML^{\perp}$) capture significant variation in portfolio returns after controlling for market risk. It is interesting to observe that the loadings on the market factor (MKT) are higher for more liquid portfolios at daily and weekly frequencies. At daily horizon, the loadings on MKT range from 0.547 for P_1 to 1.049 for P_{10}. The loadings on weekly MKT are 0.767 and 1.036 for P_1 and P_{10} respectively. Since the factor loadings show the average sensitivities of the portfolio returns to risk factors, also known as "beta", it implies that liquid portfolios are more sensitive to market risk. At each data frequency, the loadings on the $IML^{\perp}$ are highly significant. The aggregate liquidity betas decrease in portfolio liquidity and turn to negative for the liquid portfolios, indicating, low liquidity stocks bear high liquidity risk. The average risk-adjusted returns (a_p) appear to be lower for more liquid portfolios. The spreads between the returns of P_1 and P_{10} are 0.027%, 0.10%, and 0.31% at daily, weekly, and monthly frequencies respectively.

Under the rational asset pricing theory, the risk-adjusted average returns should be equal to zero. The non-zero values of individual a_p estimates in Table 1.5 imply that, the four-factor model cannot entirely explain the cross-sectional differences among the average portfolio returns. However, the liquidity factor $IML^{\perp}$ is highly significant, therefore, the four-factor specification is preferred for the rest of the analyses.

1.4.2 Innovations in Aggregate Liquidity, Volatility, and Expected Returns: Conditional Four-Factor Model

The previously shown significant impact of the lagged liquidity variables on the expected returns can be attributed to one of the two sources: (1) The innovations in liquidity and volatility are associated with the changes in portfolios' risk and return. In other words, the time variation in liquidity indirectly affects the expected returns through variation in risk. In this case, liquidity proxies for risk, and we expect the innovations and volatility to be related with the factor loadings in the conditional asset pricing model. (2) Changes in the liquidity variables do not alter the portfolios' risk, yet remain to be significant. If the *direct effect* of liquidity prevails, then the innovations in liquidity and volatility are priced because they provide non-risk related information about future prices.

Since the risk-based view suggests that the aggregate liquidity affects the expected returns through risk factors, it is useful to explore the relations between the lagged innovations, volatility, and the factors, using the following regressions:

$$F_t = \varphi_0 + \varphi_1 \hat{\eta}^*_{m,(t-1)} + \varphi_2 \hat{\sigma}_{m,(t-1)} + \nu_t \tag{1.4.1}$$

where $F_t = (MKT_t, SMB^{\perp}_t, HML^{\perp}_t, IML^{\perp}_t)'$ is the vector of four risk factors. The documented results in Table 1.6 show that, the lagged innovations in liquidity have a significant positive impact on $SMB^{\perp}$ and $IML^{\perp}$ at all data frequencies. It is also positively associated with $HML^{\perp}$ at monthly horizon. The lagged volatility measure appears to be positively related with $SMB^{\perp}$ at daily frequency. The monthly volatility, however, has a surprising negative impact on MKT and $HML^{\perp}$.

Tables 1.7 reports the least-squares estimates of the system of conditional four-factor regressions for ten liquidity-sorted portfolios. The risk-based view requires the aggregate liquidity fluctuations and volatility to be related with one or more factor loadings, i.e. $\lambda_{p1} \neq 0$ and $\lambda_{p2} \neq 0$. On the other hand, non-zero interactions between $\hat{\eta}^*_{m,(t-1)}$, $\hat{\sigma}_{m,(t-1)}$, and the intercept imply that the impact of liquidity variables on expected returns are not entirely due to time-variations in portfolios' risk. Since the focus here is the association of the factor loadings with the liquidity variables, I only report the coefficients of the interactive terms in order to reduce cluttering. Table 1.7 shows that $\hat{\eta}^*_m$ and $\hat{\sigma}_m$ are strongly associated with the time-variations in factor loadings, however, the sign and the magnitude substantially changes depending on the investment horizon.

At daily frequency (Panel A), the volatility of aggregate liquidity, $\hat{\sigma}_m$, appears to affect the loadings on $HML^{\perp}$ and $IML^{\perp}$ with an expected positive sign, although these effects are significant only for P_2, P_3, and P_9. The positive sign implies that the book-to-market and liquidity risks become more important as the volatility of aggregate liquidity increases. The significant effects of $\hat{\sigma}_m$ on the loadings of $HML^{\perp}$ and $IML^{\perp}$ are also supported by the Wald statistic values of 41.05 (p-value=0.000) and 49.73 (p-value=0.000) respectively. $\hat{\sigma}_m$ is also strongly associated with the loadings on MKT and $SMB^{\perp}$, but with a negative sign. Hence, the market and size risks of a portfolio decline with the volatility in aggregate liquidity. The decline in market risk is especially significant for more liquid portfolios. It is important to realize that the daily volatility in liquidity affects the expected portfolio returns only through its impact on the portfolios' risk. Its *direct* effect, measured by a_{p2}, is marginally and jointly equal to zero. Hence, the daily $\hat{\sigma}_m$ is a source of risk and it provides no incremental information after controlling for the time-varying factor risk.

On the other hand, the daily innovations in liquidity, $\hat{\eta}^*_m$, exhibit no association with the risk factor loadings. Its *direct* impact, measured by a_{p1}, on the expected returns prevails after controlling for the time-varying risk. The positive values of $\hat{a}_{p1}$ indicate that an unexpected increase (decrease) in aggregate liquidity results in higher (lower) expected returns the next day. Hence, at one day investment horizon, an innovation in liquidity provides information about the next day's asset prices, without affecting the risk. The Wald statistic also supports the finding for individual portfolios by rejecting the null hypothesis, $H_0 : \hat{a}_{1,1} = \hat{a}_{2,1} = ... = \hat{a}_{10,1} = 0$, with a value of 70.54 (p-value=0.00).

At the investment horizon of one week (Panel B), $\hat{\sigma}_m$ again affects the loadings on MKT and $SMB^{\perp}$ with a negative, and $HML^{\perp}$ with a positive sign. The weekly innovations in liquidity ($\hat{\eta}_m^*$) is, however, strongly associted with the size risk. The positive significant estimates of $\hat{s}_{p1}$ imply that the size risk of a portfolio increases with a positive innovation in liquidity. After controlling for the time-varying risk, the significant *direct* effect of $\hat{\eta}_m^*$ disappears. For individual portfolios, the slope coefficients of $\hat{\eta}_m^*$ are all zero, except for P_1. Also, the Wald test statistic fails to reject the null hypothesis of jointly zero $\hat{a}_{p1}$ with a value of 12.74 (p-value=0.24). Similarly, the lagged volatility of liquidity does not have any *direct* explanatory power on the expected returns. Overall, the weekly results provide supportive evidence for the risk-based view showing that the liquidity variables are associated with the changes in portfolios' risk and do not provide any non-risk related information.

The results for monthly data in Panel C also provides support for the risk-based view. The coefficients of the interactions of $\hat{\eta}_m^*$ and $\hat{\sigma}_m$ with the risk factors demonstrate that the liquidity variables capture substantial variation in factor risk. The innovations in liquidity exhibit strong negative associations with MKT and $IML^{\perp}$, implying that the market and liquidity risks decline (increase) with a positive (negative) shock to aggregate liquidity. However, $\hat{\eta}_m^*$ is again positively related with $SMB^{\perp}$, hence, an unexpected increase in liquidity results in a higher size risk. The volatility of liquidity ($\hat{\sigma}_m$) appears to be strongly correlated with all the risk factors. It affects the book-to-market risk with a negative sign. The loadings on its interactions with MKT, $SMB^{\perp}$, and $IML^{\perp}$ are positive as expected, since a higher volatility in aggregate liquidity implies higher market, size, and liquidity risk for an individual portfolio. The monthly innovations in liquidity and volatility do not have *direct* impact on the expected returns. The estimated values of $\hat{a}_{p1}$ and $\hat{a}_{p2}$ are all insignificant at individual portfolio level, except for P_2. The joint significance across portfolios is also rejected by Wald statistic with a value of 14.46 (p-value=0.15) for $\hat{a}_{p1}$, and 8.05(p-value=0.62) for $\hat{a}_{p2}$.

Overall, the results from conditional four-factor model are consistent with the risk view of liquidity at weekly and monthly frequencies. The innovations in liquidity and volatility are highly associated with loadings on the risk factors. After controlling for time-variation in factor loadings, the liquidity variables do not provide additional information about the next period's expected returns. Hence, the source of the liquidity effect on the expected returns comes from the variation in factor risk over time. The daily results are, however, different than their lower frequency counterparts. At one day investment horizon, the *direct* impact of innovations in liquidity remains to be significant after controlling for the time-varying factor risk. At this horizon, the variation in expected returns cannot be completely explained by the changes in risk and the innovations in liquidity appears to be a source of mispricing.

1.5 Conclusion

The cross-sectional relations between both the level and the risk of liquidity, and expected stock returns have long been an interest. The finding that liquidity and volatility of liquidity explain cross-sectional return differences among stocks (portfolios) implies that, at a fixed point in time liquidity based state variables convey information about returns.

Complementing the previous line of research, this study investigated the time-series relation between liquidity risk and expected returns at portfolio level. First, it addressed the question of whether liquidity based state variables provide information about the variation in expected returns. Second, it analyzed whether this information comes from a risk-based source. Using Amihud's (2002) illiquidity ratio as a proxy and Shanken's (1990) methodology, a conditional four-factor asset pricing model was employed, where the factor loadings were allowed to change over time.

Using daily, weekly, and monthly portfolio data for the period of January 1964 - December 2008, I showed that the source of the information provided by the innovations in aggregate liquidity and the volatility of aggregate liquidity vary substantially depending on the investment horizon. At weekly and monthly frequencies, the liquidity variables are highly associated with the time-variations in factor loadings, and do not provide any incremental information after accounting for these variations. At daily horizon, however, the factor loadings appear to be significantly related with volatility, but not with innovations. The innovations directly affect the expected portfolio returns without altering the risk. Hence, the impact of liquidity at daily frequency cannot be entirely attributed to changes in the risk of a portfolio.

The liquidity risk variables studied in this paper contain important information about the risk and the expected return of an asset. At daily horizon, they provide more information than just the changes in assets' risk. Since understanding the dynamics of liquidity risk help investors to make more informed investment decisions and may affect their portfolio allocations, it is worthwhile to explore the temporary dimension of the liquidity effect, and study the source(s) of the information it provides.

References

Acharya V.V. and L.H. Pedersen, (2005). Asset Pricing with Liquidity Risk, *Journal of Financial Economics* 77, 375-410.

Amihud, Y., (2002). Illiquidity and Stock Returns: Cross-Section and Time-Series Effects, *Journal of Financial Markets* 5, 31-56.

Amihud, Y. and C.M. Hurvich, (2004). Predictive Regressions: A Reduced-Bias Estimation Method, *Journal of Financial and Quantitative Analysis* 39(4), 813-841.

Amihud, Y. and H. Mendelson, (1986). Asset Pricing and the Bid-Ask Spread, *Journal of Financial Economics* 17, 223-249.

Balduzzi, P. and A. Lynch (1999). Transaction Costs and Predictability: Some utility cost calculations, *Journal of Financial Economics*, 52(1), 47-78.

Barberis, N. (2000). Investing for the Long-run when Returns are Predictable, *The Journal of Finance* 55(1), 225-64.

Bekaert G., C.R. Harvey, and C. Lundblad (2007). Liquidity and Expected Returns: Lessons from Emerging Markets, *The Review of Financial Studies* 20, 1-49.

Brennan, M.J. and A. Subrahmanyam (1996). Market Microstructure and Asset Pricing: On the Compensation for Illiquidity in Stock Returns, *Journal of Fi-*

nancial Economics 41, 441-464.

Brennan, M.J., T. Chordia, and A. Subrahmanyam (1998). Alternative Factor Specifications, Security Characteristics, and the Cross-Section of Expected Stock Returns, *Journal of Financial Economics* 49, 345-373.

Brenner, M., R. Eldor, and S. Hauser (2001). The Price of Options Illiquidity, *Journal of Finance* 56, 789-805.

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

Campbell, J.Y., S.J. Grossman, and J. Wang (1993). Trading Volume and Serial Correlation in Stock Returns, *The Quarterly Journal of Economics* 108, 905-939.

Campbell, J.Y. and S.B. Thompson (2008). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?, *The Review of Financial Studies* 21(4), 1509-1531.

Chan, J.S.P., D. Hong, and M.G. Subrahmanyam (2008). A Tale of Two Prices: Liquidity and Asset Prices in Multiple Markets, *Journal of Banking and Finance* 32(1), 947-960.

Chordia T., A. Subrahmanyam, and V.R. Anshuman (2001). Trading Activity and Expected Stock Returns, *Journal of Financial Economics* 59, 3-32.

Chordia T., R. Roll, and A. Subrahmanyam (2000). Commonality in Liquidity, *Journal of Financial Economics* 56, 3-28.

Chordia T., R. Roll, and A. Subrahmanyam (2001). Market Liquidity and Trading Activity, *Journal of Finance* 56(2), 501-530.

Chordia T., R. Roll, and A. Subrahmanyam (2008). Liquidity and Market Efficiency, *Journal of Financial Economics* 87, 249-268.

Daniel K., and D. Titman (1997). Evidence on the Characteristics of Cross-Sectional Variation in Stock Returns, *Journal of Finance* 52, 1-33.

Datar, V.T., N.Y. Naik, and R. Radcliffe (1998). Liquidity and Stock Returns: An Alternative Test, *Journal of Financial Markets* 1, 203-219.

Eckbo B.E. and O. Norli (2002). Pervasive Liquidity Risk, *Working Paper*.

Evans, G.W. and S. Honkapohja (2001). Learning and Expectations in Macroeconomics, *Princeton University Press*, Princeton, NJ.

Fama E., and K. French (1993). Common Risk Factors in the Returns on Stocks and Bonds, *Journal of Financial Economics* 33, 3-56.

Gervais, S., R. Kaniel, and D.H., Mingelgrin (2001). The High-Volume Return Premium, *Journal of Finance* 56(3), 877-919.

Goldreich, D., B. Hanke, and P. Nath (2005). The Price of Future Liquidity: Time Varying Liquidity in the U.S. Treasury Market, *Review of Finance*, 9(1), 1-32.

Hasbrouck, J. and D.J. Seppi (2001). Common Factors in Prices, Order Flows, and Liquidity, *Journal of Financial Economics* 59, 383-411.

Hasbrouck, J. (2004). Trading Costs and Returns for US Equities: The Evidence from Daily Data, *New York University.Working Paper*.

Huberman, G. and D. Halka (2001). Systematic Liquidity, *Journal of Financial Research* 24, 161-178.

Jones C.M. (2002). A Century of Stock Market Liquidity and Trading Costs, *Columbia University. Working Paper*.

Kaniel R., D. Li, and L. Starks (2005). Investor Visibility Events: Cross-Country Evidence, *Duke University. Working Paper*.

Keim D. (1983). Size-related anomalies and stock return seasonality: Further empirical evidence, *Journal of Financial Economics* 12, 13-32.

Lewellen (1999). The Time-series Relations Among Expected Return, Risk, and Book-to-Market, *Journal of Financial Economics* 54, 5-43.

Liu, W. (2006). A Liquidity-augmented Capital Asset Pricing Model, *Journal of Financial Economics* 82, 631-671.

Miralles J.L. and M.M. Miralles (2006). The Role of an Illiquidity Risk Factor in Asset Pricing: Empirical Evidence from the Spanish Stock Market, *The Quarterly Review of Economics and Finance* 46, 254-267.

Pagan, A. (1984). Econometric Issues In the Analysis of Regressions with Generated Regressors, *International Economic Review* 25(1), 221-247.

Pagan, A. (1986). Two Stage and Related Estimators and their Applications, *Review of Economic Studies* 53(4), 517-538.

Pastór L. and R.F. Stambaugh (2002). Liquidity Risk and Expected Stock Returns, *Journal of Political Economy*, 111(3), 642-685.

Pereira J.P. and H.H. Zhang (2008). Stock Returns and the Volatility of Liquidity, *Working Paper*.

Sadka R. (2006). Momentum and Post-Earnings-Announcement Drift Anomalies: The Role of Liquidity Risk, *Journal of Financial Economics* 80, 309-349.

Sadka R. and A. Scherbina (2007). Analyst Disagreement, Mispricing, and Liquidity, *Journal of Finance* 62(5), 2367-2404.

Shanken J. (1990). Intertemporal Asset Pricing: An Empirical Investigation, *Journal of Econometrics* 45, 99-120.

Silber W.L. (1991). Discounts on Restricted Stocks: The Impact of Illiquidity on Stock Prices, *Financial Analysts Journal* 47, 60-64.

Subrahmanyam A. (2008). The Implications of Liquidity and Order Flows for Neoclassical Finance, *Pacific-Basin Finance Journal* forthcoming.

Watanabe A. and M. Watanabe (2008). Time-Varying Liquidity Risk and the Cross-Section of Stock Returns, *The Review of Financial Studies* 21(6), 2449-2486.

Figure 1.1: **Time-series of Illiquidity Ratio**. Each plot shows the illiquidity ratios for three different portfolios at daily, weekly, and monthly frequencies. The bold, thin, and the dashed lines represent the illiquidity ratio of a liquid stocks portfolio (P_9), an illiquid stocks portfolio (P_2), and the market portfolio (P_m) respectively.

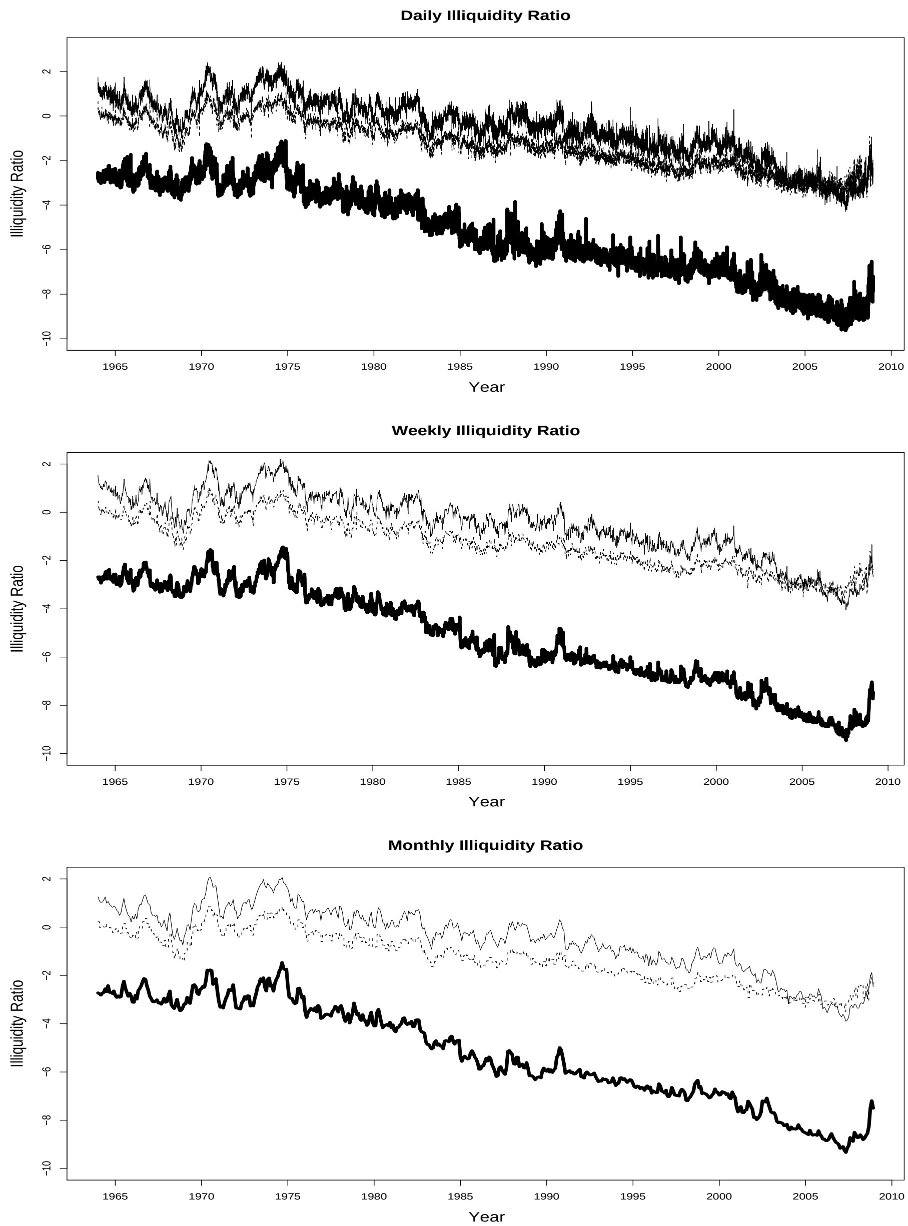
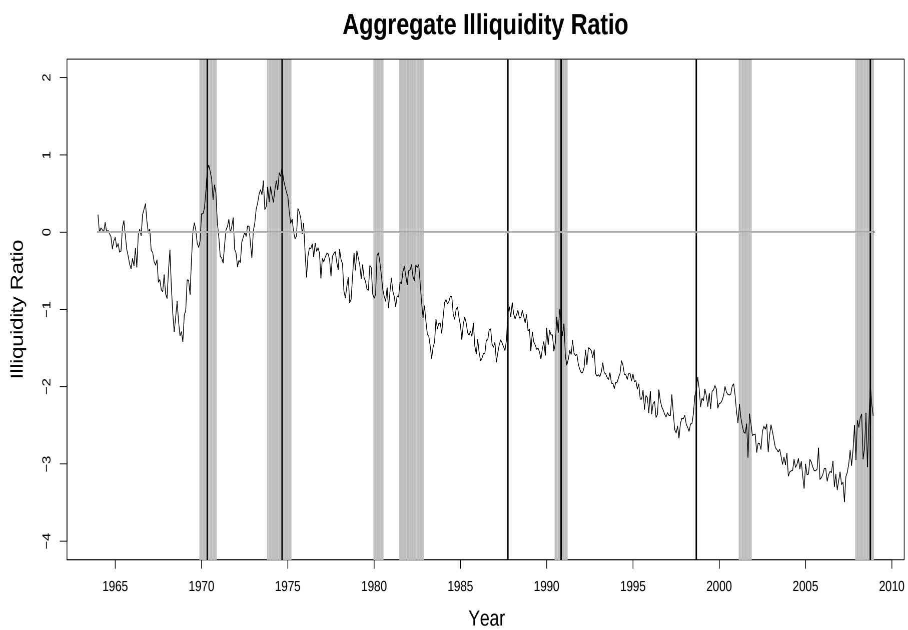


Figure 1.2: **Monthly Return and Illiquidity Ratio of the Market Portfolio**. The top figure plots the monthly aggreate illiquidity ratio and the bottom figure plots the monthly excess returns for the market portfolio P_m. The sample covers the period from January 1964 to December 2008. The gray areas represent the NBER business-cycle contraction dates. Each vertical black line corresponds to a major economic or financial event that resulted in a large decline in the aggregate liquidity.



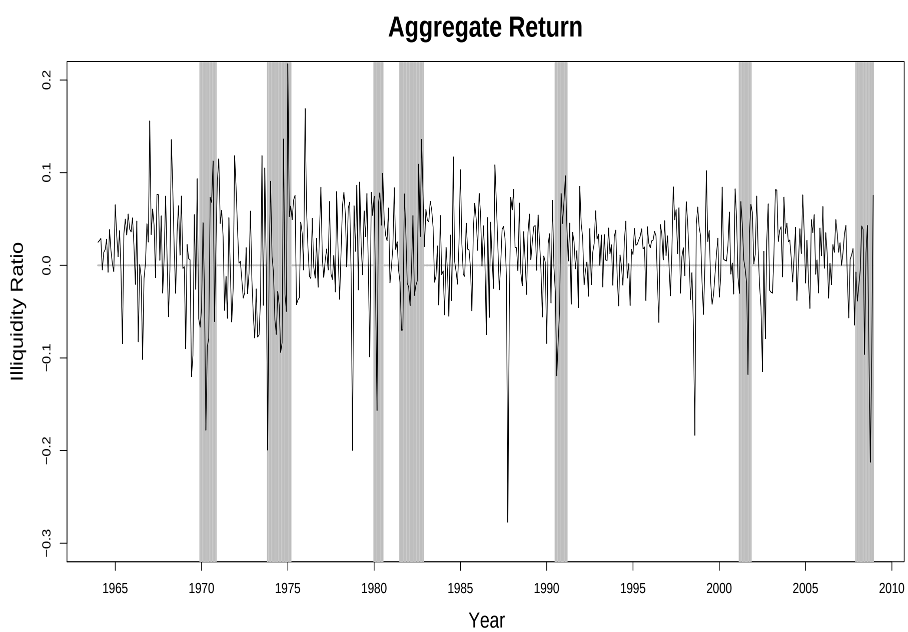


Figure 1.3: **The Innovations in Market Liquidity**. Each plot displays the innovations in aggregate liquidity, $\hat{\eta}^*_{mt}$. The innovations are the negative values of the estimated residuals from the recursive AR(1) process in Eq. (1.3.9), such that; $\hat{\eta}^*_{mt} = -(\Delta ILLIQ_{mt} - \hat{\alpha}_t - \hat{\rho}_t \Delta ILLIQ_{m,(t-1)})$.

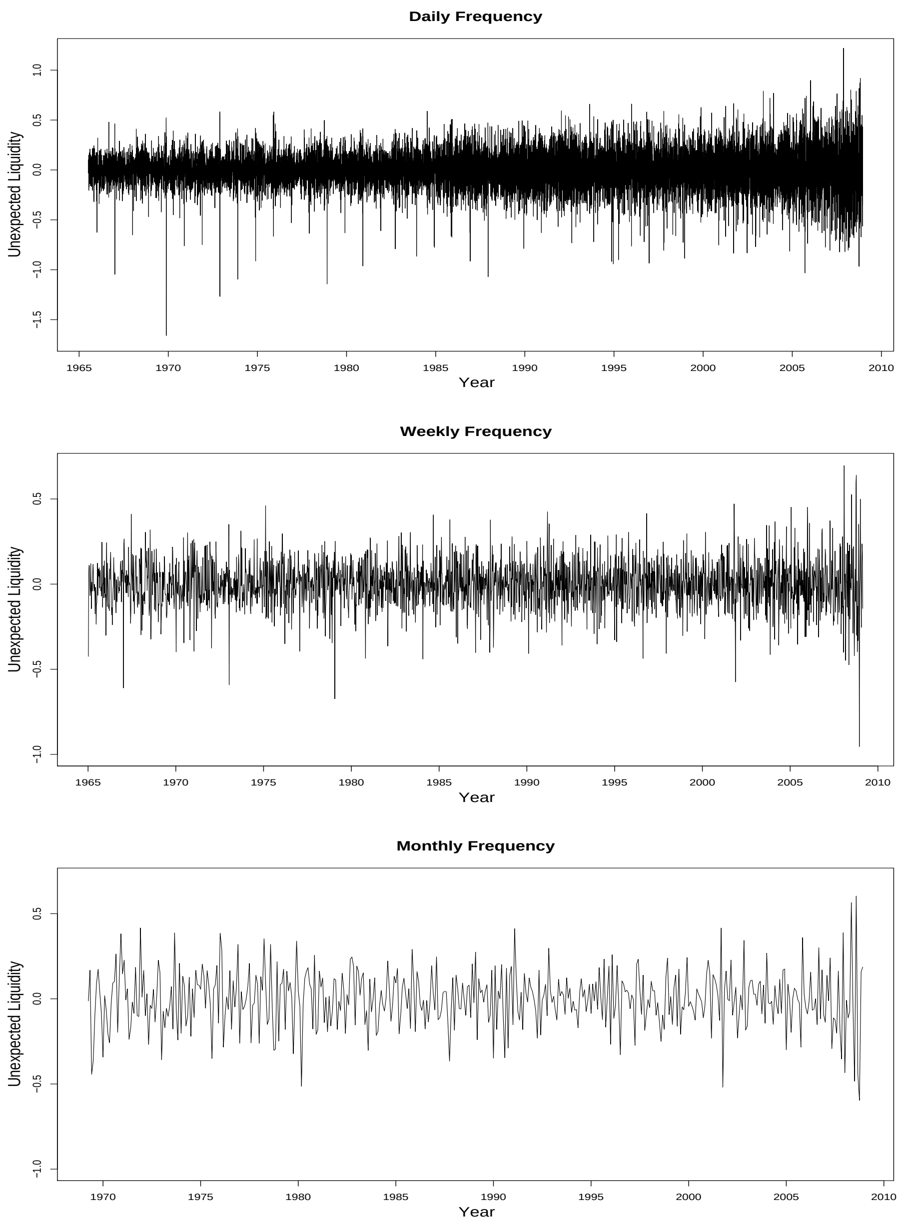


Table 1.1: **Summary statistics for the risk factors.** The means, standard deviatons, and the correlation coefficients for the standard and orthogonalized, denoted by superscript $\perp$, four risk factors are reported in the table below. The factors $SMB^{\perp}$, $HML^{\perp}$, and $IML^{\perp}$ are orthogonalized using Eq. (1.3.3) - (1.3.5).

	Mean	Stdev.	MKT	SMB	HML	IML	$SMB^{\perp}$	$HML^{\perp}$	$IML^{\perp}$
Panel A: Daily Frequency									
MKT	0.02	0.96	1.00						
SMB	0.01	0.50	-0.22	1.00					
HML	0.02	0.47	-0.43	-0.06	1.00				
IML	0.02	0.53	-0.44	0.81	0.21	1.00			
$SMB^{\perp}$	0.01	0.49	0.00	0.98	0.00	0.00	1.00		
$HML^{\perp}$	0.03	0.42	0.00	0.00	0.89	0.00	0.00	1.00	
$IML^{\perp}$	0.01	0.27	0.00	0.00	0.00	0.50	0.00	0.00	1.00
Panel B: Weekly Frequency									
MKT	0.07	2.15	1.00						
SMB	0.03	1.18	0.02	1.00					
HML	0.09	1.17	-0.42	-0.14	1.00				
IML	0.10	1.20	-0.22	0.82	0.14	1.00			
$SMB^{\perp}$	0.04	1.18	0.00	0.99	0.00	0.00	1.00		
$HML^{\perp}$	0.12	1.05	0.00	0.00	0.90	0.00	0.00	1.00	
$IML^{\perp}$	0.06	0.60	0.00	0.00	0.00	0.50	0.00	0.00	1.00
Panel C: Monthly Frequency									
MKT	0.37	4.46	1.00						
SMB	0.26	3.20	0.30	1.00					
HML	0.43	2.90	-0.38	-0.26	1.00				
IML	0.44	3.00	0.12	0.84	0.03	1.00			
$SMB^{\perp}$	0.18	3.05	0.00	0.95	0.00	0.00	1.00		
$HML^{\perp}$	0.54	2.64	0.00	0.00	0.91	0.00	0.00	1.00	
$IML^{\perp}$	0.12	1.42	0.00	0.00	0.00	0.47	0.00	0.00	1.00

Table 1.2: **Summary statistics for the portfolio returns and illiquidity ratio.** The ten liquidity-sorted portfolios are denoted by $P_1 - P_{10}$, where P_1 is the least liquid stocks portfolio and P_{10} is the most liquid stocks portfolio. P_m stands for the market portfolio, which is the portfolio of all available stocks. The percentage returns are the equally-weighted returns for each portfolio and the illiquidity ratio is Amihud's measure: $ILLIQ_{pd} = ln\left(\frac{1}{I}\sum_{i=1}^{I}\frac{|R_{id}|}{VOL_{id}}\right)$. The sample period is January 1964 - December 2008.

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_m
Panel A: Daily Frequency											
Return (%)											
Mean	0.078	0.061	0.054	0.051	0.050	0.046	0.046	0.046	0.044	0.039	0.055
Stdev	0.791	0.943	0.992	1.013	1.008	1.006	1.020	1.037	1.013	1.073	0.922
Illiquidity Ratio											
Mean	0.590	-0.492	-1.325	-1.971	-2.591	-3.172	-3.780	-4.466	-5.240	-6.350	-1.295
Stdev	0.893	1.370	1.712	1.863	1.947	2.009	2.050	2.068	2.074	2.082	1.096
Autocorr.	0.946	0.978	0.986	0.988	0.990	0.991	0.992	0.992	0.991	0.988	0.980
Adj.R^2	0.894	0.957	0.973	0.977	0.981	0.982	0.984	0.984	0.982	0.977	0.961
Panel B: Weekly Frequency											
Return (%)											
Mean	0.416	0.328	0.283	0.262	0.257	0.239	0.233	0.230	0.220	0.195	0.263
Stdev	2.264	2.427	2.440	2.462	2.456	2.422	2.431	2.436	2.310	2.309	2.302
Illiquidity Ratio											
Mean	0.606	-0.475	-1.308	-1.952	-2.573	-3.152	-3.761	-4.446	-5.218	-6.323	-1.286
Stdev	0.869	1.354	1.697	1.849	1.934	1.998	2.040	2.058	2.064	2.073	1.083
Autocorr.	0.979	0.992	0.995	0.995	0.996	0.996	0.997	0.997	0.997	0.994	0.991
Adj.R^2	0.960	0.983	0.989	0.990	0.992	0.991	0.994	0.994	0.994	0.988	0.983
Panel C: Monthly Frequency											
Return (%)											
Mean	1.817	1.433	1.242	1.150	1.129	1.049	1.023	1.011	0.964	0.854	1.151
Stdev	6.268	6.061	5.941	5.847	5.683	5.537	5.461	5.309	4.895	4.603	5.315
Illiquidity Ratio											
Mean	0.610	-0.470	-1.303	-1.946	-2.569	-3.147	-3.758	-4.443	-5.216	-6.316	-1.284
Stdev	0.861	1.349	1.694	1.847	1.932	1.996	2.038	2.057	2.062	2.070	1.078
Autocorr.	0.974	0.989	0.993	0.994	0.995	0.995	0.995	0.995	0.996	0.991	0.986
Adj.R^2	0.953	0.978	0.985	0.986	0.988	0.988	0.990	0.991	0.992	0.982	0.974

Table 1.3: **The time-series regressions of expected portfolio returns on liquidity variables**. Panel A through Panel C report the least-squares estimates for each portfolio from the system of equations, $Z_{pt} = \beta_{p0} + \beta_{p1}\hat{\eta}^*_{m,(t-1)} + \beta_{p2}\hat{\sigma}_{m,(t-1)} + \beta_{p3}Jan_t + \varepsilon_{pt}, p = 1, ..., 10$. The last two columns in the table show the average of the estimated coefficients across portfolios and the Wald statistic (W). The Wald statistic tests the null hypothesis that, an estimated coefficient is jointly equal to zero for all portfolios. It has an asymptotic χ^2 distribution under the null with d.o.f. equal to number of restrictions. The values in parentheses are the heteroskedasticity consistent standard errors. The brackets under the Wald test statistic represent the p-value for this test.

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	Avg.	W
Panel A: Daily Frequency												
const.	0.004	0.011	0.006	0.017	0.010	0.028	0.022	0.029	0.029	0.036	0.019	2.42
	(0.029)	(0.045)	(0.046)	(0.047)	(0.045)	(0.044)	(0.045)	(0.045)	(0.044)	(0.045)	(0.011)	{0.99}
$\hat{\eta}^*_{m,(t-1)}$	0.168	0.139	0.113	0.132	0.110	0.102	0.098	0.078	0.050	-0.001	0.099	35.68
	(0.047)	(0.063)	(0.065)	(0.065)	(0.062)	(0.061)	(0.061)	(0.061)	(0.059)	(0.061)	(0.048)	{0.00}
$\hat{\sigma}_{m,(t-1)}$	0.175	0.066	0.078	0.003	0.054	-0.067	-0.029	-0.064	-0.062	-0.121	0.003	1.78
	(0.159)	(0.274)	(0.284)	(0.288)	(0.276)	(0.268)	(0.276)	(0.274)	(0.266)	(0.268)	(0.037)	{1.00}
Jan_t	0.244	0.185	0.147	0.126	0.089	0.091	0.077	0.068	0.045	0.028	0.110	165.16
	(0.029)	(0.032)	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)	(0.035)	(0.066)	{0.00}
R^2	0.011	0.004	0.002	0.002	0.001	0.001	0.001	0.001	0.000	0.000		
Adj.R^2	0.011	0.003	0.002	0.002	0.001	0.001	0.000	0.000	0.000	0.000		
Panel B: Weekly Frequency												
const.	0.448	0.105	-0.020	0.034	-0.069	0.183	0.156	0.189	0.207	0.167	0.140	1.72
	(0.420)	(0.613)	(0.600)	(0.608)	(0.664)	(0.592)	(0.617)	(0.574)	(0.521)	(0.508)	(0.144)	{1.00}
$\hat{\eta}^*_{m,(t-1)}$	1.408	1.175	0.814	0.796	0.620	0.444	0.352	0.065	-0.100	-0.320	0.525	28.95
	(0.380)	(0.477)	(0.489)	(0.497)	(0.473)	(0.469)	(0.482)	(0.487)	(0.451)	(0.418)	(0.550)	{0.00}
$\hat{\sigma}_{m,(t-1)}$	-1.915	0.257	1.000	0.455	1.328	-0.817	-0.600	-0.850	-1.023	-0.835	-0.300	0.62
	(3.346)	(4.954)	(4.857)	(4.916)	(5.364)	(4.790)	(4.991)	(4.646)	(4.210)	(4.102)	(1.016)	{1.00}
Jan_t	1.139	0.833	0.692	0.615	0.472	0.494	0.426	0.374	0.282	0.179	0.551	99.05
	(0.184)	(0.200)	(0.203)	(0.203)	(0.208)	(0.201)	(0.201)	(0.197)	(0.186)	(0.188)	(0.008)	{0.00}
R^2	0.024	0.013	0.008	0.006	0.004	0.003	0.002	0.002	0.001	0.001		
Adj.R^2	0.023	0.011	0.007	0.005	0.003	0.002	0.001	0.000	0.000	0.000		
Panel C: Monthly Frequency												
const.	4.406	2.810	3.498	4.200	3.113	3.913	4.239	3.968	3.568	3.251	3.696	14.93
	(2.794)	(3.575)	(3.401)	(3.520)	(3.014)	(3.106)	(3.323)	(3.208)	(2.790)	(2.359)	(0.533)	{0.13}
$\hat{\eta}^*_{m,(t-1)}$	6.209	6.057	4.952	3.760	3.054	2.720	2.247	2.125	1.204	0.711	3.304	43.53
	(1.922)	(1.850)	(1.777)	(1.767)	(1.686)	(1.722)	(1.703)	(1.647)	(1.517)	(1.385)	(1.915)	{0.00}
$\hat{\sigma}_{m,(t-1)}$	-18.88	-11.79	-16.43	-20.86	-14.87	-19.85	-21.77	-20.20	-18.21	-17.04	-17.99	10.47
	(16.49)	(20.99)	(19.88)	(20.58)	(17.57)	(18.11)	(19.44)	(18.71)	(16.27)	(13.71)	(3.04)	{0.40}
Jan_t	-0.448	-0.142	-0.305	-0.465	-0.014	0.006	0.072	0.091	0.316	0.649	-0.024	2.62
	(0.643)	(0.649)	(0.670)	(0.681)	(0.670)	(0.656)	(0.652)	(0.665)	(0.631)	(0.608)	(0.343)	{0.99}
R^2	0.035	0.031	0.025	0.019	0.012	0.013	0.013	0.012	0.009	0.009		
Adj.R^2	0.029	0.025	0.018	0.013	0.006	0.007	0.006	0.005	0.002	0.003		

Table 1.4: **The economic significance of the innovations in liquidity**. The unconditional Sharpe ratios (S), estimated R^2 values from Eq.(1.3.14), and the R^2/S^2 ratios are provided for ten liquidity portfolios. The ratio of R^2 statistic to squared Sharpe ratio, S^2, implies the difference between the excess returns conditional on liquidity variables, and the unconditional returns. A positive value for this ratio measures the gain from the information on liquidity.

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}
Panel A: Daily Frequency										
SharpeRatio	0.071	0.040	0.032	0.028	0.027	0.024	0.023	0.022	0.021	0.016
R^2	0.011	0.004	0.002	0.002	0.001	0.001	0.001	0.001	0.000	0.000
R^2/S^2	2.211	2.462	1.931	2.622	1.392	1.782	1.890	2.008	0.000	0.000
Panel B: Weekly Frequency										
SharpeRatio	0.130	0.084	0.066	0.057	0.056	0.049	0.047	0.045	0.043	0.032
R^2	0.024	0.013	0.008	0.006	0.004	0.003	0.002	0.002	0.001	0.001
R^2/S^2	1.432	1.791	1.818	1.960	1.331	1.444	1.137	0.816	0.599	0.796
Panel C: Monthly Frequency										
SharpeRatio	0.178	0.124	0.106	0.094	0.094	0.086	0.087	0.089	0.090	0.077
R^2	0.035	0.031	0.025	0.019	0.012	0.013	0.013	0.012	0.009	0.009
R^2/S^2	1.101	2.032	2.199	2.142	1.345	1.797	1.661	1.450	1.080	1.511

Table 1.5: **Unconditional Four-Factor Model:** The system of unconditional factor models, $Z_{pt} = a_p + b_p MKT_t + s_p SMB_t^{\perp} + b_p HML_t^{\perp} + d_p IML_t^{\perp} + e_{pt}$, for $p = 1, ..., 10$, is estimated for ten liquidity portfolios. The least liquid portfolio is P_1 and the most liquid portfolio is P_{10}. The least-squares coefficient estimates are reported, with the heteroskedasticity-consistent standard deviations in parentheses.

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}
Panel A: Daily Frequency										
conts.	0.027	-0.001	-0.006	-0.008	-0.004	-0.004	-0.002	0.000	0.001	0.000
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
MKT	0.547	0.752	0.820	0.871	0.892	0.917	0.948	0.981	0.970	1.049
	(0.010)	(0.007)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.006)	(0.005)
$SMB^{\perp}$	0.580	0.769	0.817	0.757	0.671	0.527	0.381	0.233	0.018	-0.228
	(0.022)	(0.012)	(0.013)	(0.013)	(0.015)	(0.016)	(0.016)	(0.018)	(0.016)	(0.011)
$HML^{\perp}$	0.264	0.556	0.551	0.527	0.454	0.417	0.396	0.380	0.356	0.169
	(0.024)	(0.016)	(0.014)	(0.014)	(0.016)	(0.017)	(0.018)	(0.017)	(0.016)	(0.011)
$IML^{\perp}$	0.893	0.789	0.597	0.426	0.168	-0.010	-0.131	-0.286	-0.371	-0.256
	(0.029)	(0.019)	(0.020)	(0.021)	(0.022)	(0.024)	(0.022)	(0.023)	(0.020)	(0.015)
$Adj.R^2$	0.711	0.896	0.914	0.918	0.908	0.907	0.903	0.913	0.927	0.958
Panel B: Weekly Frequency										
conts.	0.108	0.004	-0.026	-0.036	-0.018	-0.024	-0.017	-0.004	0.005	0.006
	(0.020)	(0.015)	(0.015)	(0.015)	(0.016)	(0.016)	(0.016)	(0.016)	(0.014)	(0.011)
MKT	0.767	0.908	0.936	0.981	1.003	1.012	1.034	1.056	1.011	1.036
	(0.017)	(0.014)	(0.015)	(0.015)	(0.017)	(0.016)	(0.017)	(0.017)	(0.015)	(0.012)
$SMB^{\perp}$	0.802	0.862	0.831	0.755	0.650	0.514	0.381	0.235	0.027	-0.193
	(0.033)	(0.029)	(0.034)	(0.034)	(0.039)	(0.040)	(0.040)	(0.038)	(0.034)	(0.025)
$HML^{\perp}$	0.392	0.532	0.530	0.502	0.453	0.445	0.429	0.392	0.358	0.145
	(0.038)	(0.031)	(0.031)	(0.030)	(0.035)	(0.034)	(0.037)	(0.035)	(0.030)	(0.023)
$IML^{\perp}$	1.107	0.828	0.608	0.430	0.181	0.037	-0.079	-0.234	-0.324	-0.216
	(0.051)	(0.039)	(0.047)	(0.047)	(0.052)	(0.050)	(0.051)	(0.051)	(0.046)	(0.036)
$Adj.R^2$	0.825	0.918	0.915	0.921	0.906	0.907	0.904	0.911	0.918	0.947
Panel C: Monthly Frequency										
const.	0.288	-0.040	-0.159	-0.205	-0.136	-0.186	-0.161	-0.104	-0.068	-0.025
	(0.111)	(0.080)	(0.078)	(0.079)	(0.092)	(0.084)	(0.088)	(0.085)	(0.078)	(0.063)
MKT	1.039	1.081	1.090	1.121	1.110	1.107	1.113	1.096	1.018	0.976
	(0.037)	(0.021)	(0.019)	(0.020)	(0.024)	(0.020)	(0.022)	(0.022)	(0.019)	(0.018)
$SMB^{\perp}$	0.875	0.843	0.791	0.670	0.582	0.442	0.295	0.171	0.000	-0.180
	(0.053)	(0.041)	(0.051)	(0.055)	(0.059)	(0.064)	(0.067)	(0.067)	(0.055)	(0.036)
$HML^{\perp}$	0.561	0.544	0.567	0.544	0.442	0.467	0.445	0.400	0.354	0.148
	(0.056)	(0.042)	(0.042)	(0.050)	(0.054)	(0.053)	(0.054)	(0.052)	(0.046)	(0.036)
$IML^{\perp}$	1.179	0.931	0.712	0.498	0.271	0.141	0.076	-0.112	-0.225	-0.166
	(0.093)	(0.072)	(0.084)	(0.087)	(0.095)	(0.099)	(0.099)	(0.094)	(0.088)	(0.066)
$Adj.R^2$	0.884	0.931	0.931	0.928	0.897	0.902	0.897	0.897	0.904	0.934

Table 1.6: **Risk factors and liquidity variables.** The results for the regressions of the risk factors on the innovations in liquidity and volatility, $F_t = \varphi_0 + \varphi_1 \hat{\eta}^*_{m,(t-1)} + \varphi_2 \hat{\sigma}_{m,(t-1)} + \nu_t$. $F_t = (MKT_t, SMB_t^{\perp}, HML_t^{\perp}, IML_t^{\perp})'$ denotes the vector of four factors. The values in parentheses are the standard deviations.

	MKT_t	$SMB_t^{\perp}$	$HML_t^{\perp}$	$IML_t^{\perp}$
Panel A: Daily Frequency				
const.	0.031	-0.040	0.021	0.010
	(0.029)	(0.015)	(0.013)	(0.008)
$\hat{\eta}^*_{m,(t-1)}$	0.013	0.109	-0.036	0.025
	(0.049)	(0.024)	(0.021)	(0.013)
$\hat{\sigma}_{m,(t-1)}$	-0.082	0.230	0.022	0.013
	(0.160)	(0.081)	(0.069)	(0.043)
$Adj.R^2$	0.000	0.002	0.000	0.000
Panel B: Weekly Frequency				
const.	0.081	-0.190	0.010	-0.025
	(0.270)	(0.148)	(0.132)	(0.075)
$\hat{\eta}^*_{m,(t-1)}$	-0.173	0.916	0.008	0.344
	(0.326)	(0.179)	(0.159)	(0.090)
$\hat{\sigma}_{m,(t-1)}$	-0.086	1.795	0.813	0.619
	(2.092)	(1.143)	(1.021)	(0.578)
$Adj.R^2$	0.000	0.012	0.000	0.006
Panel C: Monthly Frequency				
const.	3.891	-1.403	2.565	0.304
	(1.749)	(1.168)	(1.018)	(0.537)
$\hat{\eta}^*_{m,(t-1)}$	0.815	2.379	1.588	1.487
	(1.247)	(0.833)	(0.726)	(0.382)
$\hat{\sigma}_{m,(t-1)}$	- 20.533	8.540	-11.794	-1.265
	(10.046)	(6.713)	(5.847)	(3.083)
$Adj.R^2$	0.006	0.014	0.016	0.028

Table 1.7: **Conditional Four-Factor Model:** The system of conditional factor models, $Z_{pt} = a_{pt} + b_{pt}MKT_t + s_{pt}SMB_t^{\perp} + b_{pt}HML_t^{\perp} + d_{pt}IML_t^{\perp} + e_{pt}$, for $p = 1, ..., 10$, is estimated for ten portfolios, P_1 (the least liquid) and P_{10} (the most liquid). The risk factor loadings are allowed to linearly change over time with the innovations in aggregate liquidity and volatility. The estimated least-squares coefficient values are reported below, with the heteroskedasticity-consistent standard deviations in parentheses. The p-values for the Wald test statistics, W, are reported in brackets. The superscript 1 denotes the coefficient estimate for the interactions with the lagged liquidity innovations $\hat{\eta}^*_{m,(t-1)}$. The estimates for the interactions with the standard deviation of liquidity $\hat{\sigma}_{m,(t-1)}$ are denoted by the superscript 2.

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	W
Panel A: Daily Frequency											
$const.^1$	0.061	0.043	0.018	0.042	0.028	0.040	0.054	0.053	0.057	0.025	70.54
	(0.024)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.017)	(0.017)	(0.015)	(0.011)	{0.00}
$const.^2$	0.175	-0.023	-0.009	-0.058	0.017	-0.074	-0.007	-0.005	0.011	-0.012	7.33
	(0.088)	(0.059)	(0.056)	(0.054)	(0.055)	(0.056)	(0.062)	(0.058)	(0.053)	(0.039)	{0.69}
MKT^1	0.005	0.015	-0.010	-0.006	-0.017	0.001	-0.003	0.014	-0.013	0.007	1.36
	(0.034)	(0.028)	(0.030)	(0.029)	(0.028)	(0.029)	(0.032)	(0.032)	(0.030)	(0.020)	{1.00}
MKT^2	-1.630	-0.062	0.121	0.124	-0.009	-0.202	-0.204	-0.286	-0.331	-0.397	236.67
	(0.126)	(0.097)	(0.079)	(0.085)	(0.088)	(0.082)	(0.102)	(0.092)	(0.102)	(0.067)	{0.00}
$SMB^{\perp 1}$	0.005	-0.057	-0.061	0.024	-0.009	-0.023	0.032	-0.051	-0.060	0.023	5.48
	(0.089)	(0.043)	(0.054)	(0.040)	(0.048)	(0.056)	(0.058)	(0.078)	(0.066)	(0.037)	{0.86}
$SMB^{\perp 2}$	-2.345	0.042	0.517	0.488	-0.264	-0.552	-0.643	-0.604	-0.593	-0.126	110.27
	(0.324)	(0.178)	(0.142)	(0.153)	(0.180)	(0.205)	(0.239)	(0.235)	(0.188)	(0.122)	{0.00}
$HML^{\perp 1}$	0.051	0.067	0.038	0.052	0.086	-0.002	0.104	0.093	0.076	-0.025	9.38
	(0.084)	(0.063)	(0.068)	(0.062)	(0.065)	(0.066)	(0.076)	(0.074)	(0.067)	(0.043)	{0.50}
$HML^{\perp 2}$	-0.051	0.786	0.281	0.205	0.095	0.148	0.197	0.294	0.235	0.716	41.05
	(0.297)	(0.226)	(0.193)	(0.187)	(0.205)	(0.219)	(0.238)	(0.232)	(0.199)	(0.155)	{0.00}
$IML^{\perp 1}$	-0.164	-0.125	-0.091	-0.001	0.022	-0.015	-0.090	-0.080	-0.158	-0.030	13.07
	(0.123)	(0.074)	(0.072)	(0.075)	(0.081)	(0.081)	(0.083)	(0.087)	(0.076)	(0.051)	{0.22}
$IML^{\perp 2}$	-0.679	1.129	0.488	0.000	-0.421	-0.350	-0.081	-0.402	0.498	0.855	49.73
	(0.482)	(0.308)	(0.227)	(0.240)	(0.268)	(0.308)	(0.353)	(0.349)	(0.286)	(0.184)	{0.00}
$Adj.R^2$	0.752	0.898	0.915	0.918	0.908	0.907	0.904	0.914	0.928	0.959	

Table 1.7: Conditional Four-Factor Model (contd.)

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	W
Panel B: Weekly Frequency											
$const.^1$	0.301	0.148	-0.047	0.075	0.096	0.086	0.129	0.071	0.132	0.110	12.74
	(0.156)	(0.111)	(0.113)	(0.112)	(0.107)	(0.110)	(0.107)	(0.114)	(0.096)	(0.091)	{0.24}
$const.^2$	-0.293	-0.237	0.352	-0.111	0.663	-1.008	-0.709	-0.305	-0.149	-0.086	4.94
	(1.421)	(0.747)	(0.720)	(0.748)	(0.751)	(0.646)	(0.702)	(0.660)	(0.604)	(0.531)	{0.90}
MKT^1	0.060	0.012	0.000	-0.025	-0.009	-0.039	-0.147	-0.069	-0.028	0.062	8.74
	(0.086)	(0.074)	(0.069)	(0.072)	(0.066)	(0.063)	(0.063)	(0.075)	(0.064)	(0.057)	{0.56}
MKT^2	-1.840	0.217	0.162	-0.069	0.480	-0.372	0.026	-0.552	-0.463	-0.576	18.99
	(0.608)	(0.402)	(0.366)	(0.351)	(0.479)	(0.345)	(0.410)	(0.429)	(0.367)	(0.292)	{0.04}
$SMB^{\perp 1}$	0.149	0.208	0.320	0.421	0.337	0.370	0.324	0.388	0.218	0.265	40.84
	(0.180)	(0.145)	(0.148)	(0.150)	(0.165)	(0.169)	(0.168)	(0.164)	(0.147)	(0.118)	{0.00}
$SMB^{\perp 2}$	-3.208	-1.479	-0.783	0.362	-0.035	-0.797	-0.911	-2.218	-1.142	-0.737	18.71
	(1.510)	(0.870)	(0.964)	(0.954)	(1.008)	(0.948)	(0.983)	(0.890)	(0.880)	(0.722)	{0.04}
$HML^{\perp 1}$	0.139	-0.041	-0.113	0.039	-0.119	-0.033	-0.102	0.057	-0.039	0.018	2.01
	(0.209)	(0.164)	(0.165)	(0.170)	(0.177)	(0.180)	(0.177)	(0.190)	(0.166)	(0.145)	{1.00}
$HML^{\perp 2}$	-0.159	1.453	0.980	0.739	0.785	1.677	1.440	1.728	0.327	1.798	24.64
	(1.453)	(0.898)	(0.754)	(0.827)	(0.933)	(0.875)	(0.952)	(0.910)	(0.757)	(0.598)	{0.01}
$IML^{\perp 1}$	0.110	0.317	-0.188	-0.119	-0.086	-0.005	0.208	0.029	-0.080	0.017	5.08
	(0.293)	(0.205)	(0.209)	(0.197)	(0.210)	(0.213)	(0.209)	(0.227)	(0.176)	(0.179)	{0.89}
$IML^{\perp 2}$	-3.283	-1.605	-1.424	-1.875	-3.423	-0.857	-1.542	-1.928	-1.422	-1.275	20.81
	(2.296)	(1.223)	(1.221)	(1.324)	(1.447)	(1.224)	(1.229)	(1.162)	(1.136)	(0.980)	{0.02}
$Adj.R^2$	0.832	0.919	0.916	0.922	0.908	0.908	0.905	0.913	0.918	0.948	

Table 1.7: Conditional Four-Factor Model (contd.)

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	W
Panel C: Monthly Frequency											
$const.^1$	0.716	1.125	0.153	-0.342	-0.130	-0.102	-0.088	0.494	0.228	0.389	14.46
	(0.591)	(0.386)	(0.418)	(0.403)	(0.437)	(0.460)	(0.442)	(0.419)	(0.356)	(0.303)	{0.15}
$const.^2$	-6.195	8.440	2.410	4.065	2.782	0.809	-0.910	1.300	0.963	-0.287	8.05
	(5.306)	(4.149)	(3.410)	(3.583)	(4.216)	(3.652)	(4.031)	(3.692)	(3.181)	(2.990)	{0.62}
MKT^1	0.060	-0.014	-0.385	-0.164	-0.308	-0.266	-0.209	-0.258	-0.157	-0.126	48.31
	(0.152)	(0.077)	(0.125)	(0.111)	(0.099)	(0.098)	(0.097)	(0.089)	(0.076)	(0.091)	{0.00}
MKT^2	3.346	4.884	2.135	3.860	3.539	3.309	4.829	4.104	4.445	2.541	316.25
	(1.493)	(0.688)	(0.780)	(0.883)	(0.702)	(0.618)	(0.678)	(0.640)	(0.520)	(0.648)	{0.00}
$SMB^{\perp 1}$	0.556	0.321	0.155	0.379	0.156	0.310	0.367	0.333	0.393	0.307	30.89
	(0.249)	(0.174)	(0.193)	(0.201)	(0.218)	(0.205)	(0.207)	(0.201)	(0.178)	(0.142)	{0.00}
$SMB^{\perp 2}$	9.834	7.102	8.782	7.646	9.577	9.735	10.240	9.879	9.327	6.981	319.69
	(2.128)	(1.597)	(1.557)	(1.654)	(1.719)	(1.560)	(1.602)	(1.603)	(1.383)	(1.254)	{0.00}
$HML^{\perp 1}$	-0.077	0.070	-0.028	-0.206	-0.110	-0.291	-0.099	-0.256	-0.066	-0.066	5.49
	(0.242)	(0.175)	(0.201)	(0.205)	(0.214)	(0.226)	(0.220)	(0.197)	(0.171)	(0.136)	{0.86}
$HML^{\perp 2}$	-13.286	-10.978	-8.456	-10.797	-14.783	-12.268	-12.980	-12.798	-12.972	-9.089	542.50
	(2.186)	(1.665)	(1.556)	(1.615)	(1.875)	(1.576)	(1.703)	(1.521)	(1.340)	(1.394)	{0.00}
$IML^{\perp 1}$	-0.134	-0.822	-0.114	-0.585	-0.640	-0.546	-0.721	-0.611	-0.621	-0.399	35.28
	(0.414)	(0.282)	(0.303)	(0.323)	(0.324)	(0.340)	(0.336)	(0.301)	(0.296)	(0.209)	{0.00}
$IML^{\perp 2}$	6.575	3.059	5.092	10.232	9.490	8.629	4.201	6.499	5.337	3.313	45.90
	(4.354)	(3.191)	(3.131)	(3.307)	(3.116)	(3.271)	(2.852)	(2.790)	(2.416)	(2.485)	{0.00}
$Adj.R^2$	0.906	0.950	0.943	0.943	0.924	0.923	0.925	0.922	0.936	0.951	

Chapter 2

Exact Distribution-Free Tests of Mean-Variance Efficiency

Abstract

This paper develops exact distribution-free tests of unconditional mean-variance efficiency. These new tests allow for unknown forms of non-normalities, conditional heteroskedasticity, and other non-linear temporal dependencies among the absolute values of the error terms in the asset pricing model. Exactness here rests on the assumption that the joint temporal error density is symmetric around zero. This still leaves open the possibility of return distribution asymmetry via coskewness with the benchmark portfolio. A simulation study shows that the new tests have very good power relative to that of many commonly used tests. The inference procedures developed are further illustrated by tests of the mean-variance efficiency of a market index using a forty-two-year sample of monthly returns on ten U.S. equity portfolios.

JEL classification: C12; C14; C22; G11; G12

Keywords: CAPM; Conditional heteroskedasticity; Non-parametric tests; Robust inference

2.1 Introduction

The celebrated capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965) extends the notion of a mean-variance efficient portfolio to the portfolio of all invested wealth—the market portfolio. A given portfolio is mean-variance efficient if it has the smallest possible variance of return given its expected return, or, more appropriately, if it has the largest expected return given its variance. This theory implies that expected excess returns on assets are linearly related to the slope, or beta, of their regression on the expected excess return of the benchmark portfolio. Here excess returns are those in excess of the riskless rate of return. Under mean-variance efficiency, the risk premium of an asset is a linear function of the asset's beta.

Empirical tests of the mean-variance efficiency hypothesis are usually conducted within the context of a multivariate linear regression. The application of tests based on asymptotic theory can lead to misleading conclusions as the approximation to the finite-sample distribution of test statistics can be quite poor, especially as the number of equations included in the system increases; see Shanken (1996), Campbell, Lo, MacKinlay (1997), Dufour and Khalaf (2002), and the numerical evidence presented here. The findings show that many standard parametric tests are unreliable, rejecting the null hypothesis of mean-variance efficiency far too often.

Gibbons, Ross, and Shanken (1989) (GRS) propose a truly finite-sample test. The exact distribution theory for their multivariate F-test rests on the assumption that the regression error terms are independent over time and jointly normally distributed each period, conditional on the returns of the benchmark portfolio under test. These assumptions are at odds with some well-known facts about financial asset returns. Indeed, it has long been recognized that financial returns depart from Gaussian conditions; see Fama (1965), Blattberg and Gonedes (1974), and Hsu (1982). In particular, the distribution of asset returns appears to have fatter tails than those of a normal distribution.

Beaulieu, Dufour, and Khalaf (2007) (BDK) propose an exact randomized likelihood-based test procedure that relaxes normality. Their framework assumes that the error distribution is known, or at least specified up to some unknown nuisance parameters. If normality is assumed, the BDK test becomes the simulation-based equivalent of the GRS test. More generally, the BDK test procedure can be thought of as an exact parametric bootstrap. When nuisance parameters are present, the computational cost of the BDK test procedure grows with the number of nuisance parameters since it then involves finding maximal p-values over a confidence region for the intervening nuisance parameters. The first-step confidence region is established by inverting a simulation-based goodness-of-fit test (of the assumed distribution), which involves a grid search over the nuisance parameter space.

Without such parametric assumptions it would seem difficult to derive an exact finite-sample distribution theory. Despite this apparent difficulty, we propose in this paper new non-randomized tests that are exact in finite samples without any parametric assumptions about the distribution of the error terms in the simple multivariate linear regression model. The methods exploit results derived by Luger (2003) in the context of testing for a random walk. Here we propose three testing approaches for joint inference on several parameters that differ mainly by what is assumed about the covariance of the errors across equations.

The first approach is an induced test procedure that allows for arbitrary covariances in the cross-section of error terms. The price to pay for this extra flexibility is that those tests can be conservative and lead to power losses if the number of test assets is large. The second approach assumes that the errors are independent across equations, conditional on the returns of the benchmark portfolio. This delivers tests with the correct size and more power than the induced tests. The third approach is based on a simple linear combination of the test assets (or portfolios of test assets) and, like the second approach, provides a test procedure with the correct size no matter the number of included assets. These single-portfolio tests allow some forms of covariation in the cross-section of error terms. The number of assets in the cross-section may even exceed the number of time-series observations, making these tests particularly attractive when testing mean-variance efficiency with many test assets or when the portfolios have relatively short histories. This stands in contrast to extant approaches based on estimates of the covariance matrix of the regression errors. In order to avoid singularities, those approaches require the size of the cross-section be less than that of the time series. The distribution-free approaches to inference developed here do not require the error covariance matrix.

The proposed distribution-free (or non-parametric) tests have several appealing features, since they are built on the mere assumption that the joint temporal error density is symmetric around zero. This means that no restrictions are placed on the degree of non-normality or the degree of heterogeneity across marginal distributions. In fact, the existence of moments need not be assumed for the validity of the new tests. It is important to note that this framework still leaves open the possibility of asymmetries in the distribution of test asset returns via coskewness with the benchmark portfolio.

Asset returns typically display clear patterns of volatility clustering for which generalized autoregressive conditional heteroskedasticity (GARCH) models are often used; see Bollerslev (1986). The tests proposed here allow not only for non-normalities, but also for *unknown* forms of conditional heteroskedasticity and other intertemporal dependencies among the absolute values of the error terms in the asset pricing model. Such forms of intertemporal dependencies invalidate the exact statistical theory underlying the parametric GRS and BDK test procedures. Since it is well known that asset returns depart from homogeneous conditions, the new tests of the mean-variance efficiency hypothesis developed here offer a valid and useful distribution-free testing alternative to potentially misleading parametric procedures.

Section 2 of the paper presents the framework, the hypothesis of interest, and the extant GRS and BDK test procedures that are exact under parametric distributional assumptions. Section 3 describes the building blocks of our three approaches to distribution-free inference. Section 4 is divided into three subsections, each one describing a proposed test procedure. Section 5 begins by presenting some additional (asymptotic) extant tests of mean-variance efficiency and then presents the results of some simulation examples to illustrate the behavior of the proposed tests relative to the commonly used procedures. Section 6 provides an empirical illustration of the new tests in the context of the Sharpe-Lintner version of the CAPM. Section 7 concludes.

2.2 Exact parametric tests

A benchmark portfolio with excess returns r_p is said to be mean-variance efficient with respect to a given set of N test assets with excess returns r_i, $i = 1, ..., N$, if it is not possible to form another portfolio of those N assets and the benchmark portfolio with the same expected return as r_p but a lower variance, or equivalently, with the same variance but a higher expected return. More formally, portfolio p is mean-variance efficient if the following first-order condition is satisfied for the N test assets:

$$E[r_{it}] = \beta_i E[r_{pt}], \quad i = 1, ..., N, \tag{2.2.1}$$

where r_{it} and r_{pt} are the time-t returns on asset i and portfolio p, respectively, in excess of the riskless rate of return. The term β_i captures the degree of association between the expected excess return on the individual asset i and the expected excess return for portfolio p. Accordingly, assets with higher betas should offer in equilibrium higher expected returns. Consider the excess-return system of equations

$$r_{it} = a_i + \beta_i r_{pt} + \varepsilon_{it}, \quad t = 1, ..., T, \ i = 1, ..., N, \tag{2.2.2}$$

where ε_{it} is a random error term for asset i in period t with the property that $E[\varepsilon_{it}] = 0$. The specification in (2.2.2) is a seemingly unrelated equations model. The mean-variance efficiency condition in (2.2.1) can then be assessed by testing

$$H_0 : a_i = 0, \quad i = 1, ..., N, \tag{2.2.3}$$

in the context of (2.2.2). This null hypothesis follows from a comparison of the unconditional expectation of (2.2.2) to the mean-variance efficiency condition in (2.2.1). If H_0 does not hold, it would be possible to obtain a higher expected return with no higher risk, contradicting the hypothesis that portfolio p is mean-variance efficient.

The exact Wald test of H_0 in (2.2.3) proposed by GRS assumes that the error terms $(\varepsilon_{1t}, ..., \varepsilon_{Nt})$ are jointly normally distributed around zero each period, conditional on the excess returns $(r_{p1}, ..., r_{pT})$; i.e., the regressor variable is assumed to be strictly exogenous. Further, the error terms are assumed independent over time. Under normality, the methods of maximum likelihood and ordinary least squares yield the same estimates of the a_i's and the β_i's in (2.2.2). Collect those estimates in $\hat{a} = (\hat{a}_1, ..., \hat{a}_N)'$ and $\hat{\beta} = (\hat{\beta}_1, ..., \hat{\beta}_N)'$, and let $\hat{\Sigma}$ denote the unconstrained maximum-likelihood estimate of Σ—the $N \times N$ error covariance matrix. The GRS test statistic is

$$J_1 = \frac{(T-N-1)}{N}\left[1 + \frac{\hat{\mu}_p^2}{\hat{\sigma}_p^2}\right]^{-1} \hat{a}'\hat{\Sigma}^{-1}\hat{a}, \tag{2.2.4}$$

where $\hat{\mu}_p$ is the sample mean of r_{pt} and $\hat{\sigma}_p^2$ is the maximum-likelihood estimate of the variance of r_{pt}. Under the null hypothesis H_0, the statistic J_1 follows an exact central F distribution with N degrees of freedom in the numerator and $(T-N-1)$ degrees of freedom in the denominator.

BDK also derive exact tests of mean-variance efficiency. In the context of (2.2.2), they assume the errors have the following distribution structure:

$$(\varepsilon_{1t}, ..., \varepsilon_{Nt})' = \Lambda U_t, \quad t = 1, ..., T, \tag{2.2.5}$$

where Λ is an unknown, non-singular matrix and the distribution of the U_t's is either completely known or specified up to a finite number of nuisance parameters, ν. BDK propose Monte Carlo (MC) tests based on the likelihood ratio (LR) statistic

$$J_2 = T\left(\log|\hat{\Sigma}_0| - \log|\hat{\Sigma}|\right), \tag{2.2.6}$$

where $\hat{\Sigma}_0$ is the constrained estimator of Σ. The key property for the MC tests is that (2.2.6) is invariant with respect to Λ. When U_t is Gaussian, the BDK test is the simulation-based equivalent of the GRS test since J_1 and J_2 are related via the monotonic transformation $J_1 = \frac{T-N-1}{N}(\exp[J_2/T] - 1)$; see Equation (5.3.40) in Campbell, Lo, and MacKinlay (1997). More generally, size-correct MC tests of H_0 can be performed for any assumed distribution of the U_t's that does not involve nuisance parameters. BDK also propose a maximized MC (MMC) procedure for situations with an incompletely specified error distribution, such as the multivariate Student-t distribution with unknown degrees of freedom. The MMC procedure consists of two steps. First, an exact confidence set for the nuisance parameters is constructed by inverting an MC goodness-of-fit test of the hypothesis in (2.2.5) with ν specified. Inverting that test involves assembling by grid search the values of ν that are not rejected at a specific level of significance, say α_1. In the second step, the p-value of the MC LR test is maximized over the first-step confidence set. If the resulting MMC p-value is compared to a cutoff level α_2, then the overall level of this (computationally intensive) procedure is $\alpha = \alpha_1 + \alpha_2$; see BDK for details and Dufour and Khalaf (2001) for a more general discussion about the technique of MC testing. Like any MC test procedure, the BDK approach is exact only if the maintained distribution structure actually holds.

2.3 Building blocks

Within the context of model (2.2.2), we develop distribution-free tests of the mean-variance efficiency null hypothesis in (2.2.3). Elliptically symmetric distributions of returns are very attractive in this context because they guarantee that mean-variance analysis is fully compatible with expected utility maximization; see Chamberlain (1983), Owen and Rabinovitch (1983), and Berk (1997). If a random vector of returns follows an elliptically symmetric distribution, then the marginal distribution of any component of that vector is also elliptically symmetric; see Ingersoll (1987, p. 104). Consistent with that fact, the proposed tests are based on an assumption of symmetry for the marginal error distributions and proceed by taking long differences between return observations separated by $m = T/2$ periods. Here we assume that T is even so that the midpoint m is an integer. The approach is based on Luger (2003) and makes the following assumptions about the error terms for test asset i and the excess returns of portfolio p:

Assumption 1. The density of $(\varepsilon_{i1}, ..., \varepsilon_{iT})$ is symmetric around zero, conditional on $R_p = (r_{p1}, ..., r_{pT})'$. Further, $\Pr[r_{pt} = 0] = 0$ for $t = 1, ..., T$.

Assumption 2. The random variables ε_{it}, $t = 1, ..., T$, are continuous, conditional on R_p.

The approach proposed here is based on distribution-free tests about the medians of ε_i, $i = 1, ..., N$. Under Assumption 1, these become equivalent to the usual tests (e.g. the GRS test) that maintain $E[\varepsilon_{it} \mid R_p] = 0$. Note that the distribution of r_{pt} may be skewed thereby inducing asymmetry in the unconditional distribution of r_{it}. In that case, a given portfolio can still be mean-variance efficient provided that investors' expected utility depends only on the mean and variance of a portfolio's return.

What distinguishes the present framework from Luger (2003) is that here the multivariate symmetry assumption is made conditional on $R_p = (r_{p1}, ..., r_{pT})'$. It should be noted that this conditioning on R_p is entirely consistent with arbitrage pricing theory. It is also a common approach used by GRS and BDK, for instance, to obtain a finite-sample distribution theory. The distributional assumptions made by those authors, however, are far more restrictive than Assumptions 1 and 2. GRS require independent and jointly normal errors, while BDK relax the normality assumption but nevertheless require that the error distribution structure be specified up to a finite number of nuisance parameters. Here the conditional distribution of $(\varepsilon_{i1}, ..., \varepsilon_{iT})$ given R_p is completely unspecified. This means that no restrictions are placed on the degree of non-normality or the degree of heterogeneity across marginal distributions. It should be noted that several popular models of time-varying conditional variance, such as GARCH-type or stochastic volatility models with continuous and symmetric innovations, satisfy Assumptions 1 and 2. In fact, those assumptions allow for the presence of *unknown* forms of conditional heteroskedasticity (or any other form of non-linear dependence among the $|\varepsilon_{it}|$'s). Furthermore, the conditional variances need not be finite nor even follow a stationary process, and moments need not exist.

Define the scaled returns r_{it}/r_{pt}, for $i = 1, ..., N$ and $t = 1, ..., T$. Under the statistical specification in (2.2.2), these scaled returns can be expressed as

$$\frac{r_{it}}{r_{pt}} = \frac{a_i}{r_{pt}} + \beta_i + \frac{\varepsilon_{it}}{r_{pt}},$$

thereby relegating the nuisance parameter β_i to the role of intercept. Following Theil (1971, p. 616), that parameter can be eliminated from the inference problem via the long differences

$$z_{it} = d_t^m \times x_{pt}, \tag{2.3.1}$$

for $t = 1, ..., m$, where

$$d_t^m = \left(\frac{r_{i,t+m}}{r_{p,t+m}} - \frac{r_{it}}{r_{pt}} \right) \text{ and } x_{pt} = \frac{(r_{pt} - r_{p,t+m})}{r_{pt} r_{p,t+m}}.$$

The quantity in (2.3.1) is the building block of the distribution-free methods proposed here. An important aspect of the sequence of long differences d_t^m, $t = 1, ..., m$, is that it contains no overlaps. The methodology of Luger (2003) yields exact tests when built on d_t^m. The multiplication of d_t^m by x_{pt} seen in (2.3.1) is done in order to improve power. To see why this is, note that

$$d_t^m = a_i x_{pt} + \left(\frac{\varepsilon_{i,t+m}}{r_{p,t+m}} - \frac{\varepsilon_{it}}{r_{pt}} \right). \tag{2.3.2}$$

The proof of Proposition 1 below shows that $(\varepsilon_{i,t+m}/r_{p,t+m} - \varepsilon_{it}/r_{pt})$ has a symmetric distribution. Suppose that $x_{pt} = (1/r_{p,t+m} - 1/r_{pt})$ is also symmetrically distributed, independently of the ε_{it}'s. Then the quantity in (2.3.2) would have median zero no matter the value of a_i, so that any sign statistic based on that quantity would have nothing but trivial power. Multiplying d_t^m with x_{pt} to get z_{it} shifts its median to the right or left, depending on the sign and magnitude of a_i, and hence delivers sign tests with power to detect $a_i \neq 0$. While such a modification results in only approximately exact tests in Luger's (2003) random walk context, here the tests are based on truly exact finite-sample pivots owing to the fact that we can condition on R_p.

Define a sign function as $s[z] = 1$ if $z > 0$, and $s[z] = 0$ if $z \leq 0$. The following two propositions are adapted from Luger (2003) and proofs are provided in the appendix for the sake of self-containment. Note that when $a_i = 0$, the statistic z_{it} in (2.3.1) is a function only of $(\varepsilon_{i,t+m}/r_{p,t+m} - \varepsilon_{it}/r_{pt})x_{pt}$.

Proposition 1. *Let $\varepsilon_{i1}, ..., \varepsilon_{iT}$ be a sequence of random variables that satisfies Assumptions 1 and 2. Then, the random variable $s[(\varepsilon_{i,t+m}/r_{p,t+m} - \varepsilon_{it}/r_{pt})x_{pt}]$ is distributed like a Bernoulli variable B_t, such that $\Pr[B_t = 1] = \Pr[B_t = 0] = 1/2$ for $t = 1, 2, ..., m$.*

This result shows that the random variables $s[z_{it}]$, $t = 1, ..., m$, are identically distributed under the null hypothesis. The next proposition shows that those random variables are also mutually independent, thereby paving the way for test procedures based on a class of linear signed rank statistics defined by

$$SR_i = \sum_{t=1}^{m} s[z_{it}]\varphi[Rank(|z_{it}|)], \tag{2.3.3}$$

where $Rank(|z_{it}|)$ is the rank of $|z_{it}|$ when $|z_{i1}|, ..., |z_{im}|$ are placed in ascending order of magnitude, and $0 \leq \varphi[1] \leq ... \leq \varphi[m]$ is a set of scores with $\varphi[m] > 0$. In the following, the symbol $\stackrel{d}{=}$ stands for the equality in distribution.

Proposition 2. *Let $\varepsilon_{i1}, ..., \varepsilon_{iT}$ be a sequence of random variables that satisfies Assumptions 1 and 2. Then, the null distribution of any linear signed rank statistic defined by (2.3.3) has the property that*

$$SR_i \stackrel{d}{=} \sum_{j=1}^{m} B_j \varphi[j],$$

where $B_1, ..., B_m$ are mutually independent uniform Bernoulli variables on $\{0, 1\}$.

The choice of score function yields different linear signed rank statistics. The two statistics of greatest interest are the sign statistic, obtained from the score function $\varphi[j] = 1$:

$$S_i = \sum_{t=1}^{m} s[z_{it}],$$

and the Wilcoxon signed rank statistic

$$W_i = \sum_{t=1}^{m} s[z_{it}]Rank(|z_{it}|),$$

obtained with $\varphi[j] = j$. The sign statistic follows a binomial distribution with number of trials m and probability of success $1/2$. The distribution of the Wilcoxon statistic has been tabulated for various values of m; see Table A.4 in Hollander and Wolfe (1973) for $m \leq 15$. For larger values, the standard normal distribution provides a very good approximation to the standardized variate. The same is true of the standardized sign statistic. The basic result (found in Randles and Wolfe 1979, Section 10.2) is that the standardized sign and Wilcoxon signed rank statistics

$$S_i^* = \frac{S_i - m/2}{\sqrt{m/4}} \text{ and } W_i^* = \frac{W_i - m(m+1)/4}{\sqrt{m(m+1)(2m+1)/24}}, \tag{2.3.4}$$

where null means and variances are used, have limiting (as $m \to \infty$) standard normal distributions. It should be noted that the asymptotic approximation works extremely well even for small values of m since the Binomial and Wilcoxon distributions are close to normal.

In principle, the approach adopted here could be extended to allow for additional covariates. For example, suppose returns were represented by

$$r_{it} = a_i + \beta_i r_{pt} + \gamma_i r_{ft} + \varepsilon_{it}, \quad t = 1, ..., T, \; i = 1, ..., N,$$

where r_{ft} is a second common factor. In that case, the basic building of the distribution-free tests would become

$$z_{it} = \left(\frac{d_{t+m/2}^m}{x_{f,t+m/2}} - \frac{d_t^m}{x_{ft}} \right) \left(\frac{x_{p,t+m/2}}{x_{f,t+m/2}} - \frac{x_{pt}}{x_{ft}} \right),$$

where $x_{ft} = (r_{pt} r_{f,t+m} - r_{ft} r_{p,t+m}) / r_{pt} r_{p,t+m}$ for $t = 1, ..., m/2$.

2.4 Exact distribution-free tests

2.4.1 Sup-type tests

A test of H_0 with a single test asset can be performed simply by comparing either of the test statistics in (2.3.4) with the appropriate critical values from the standard normal distribution. In order to test the null hypothesis with several test assets indexed by $i = 1, ..., N$, consider the test statistics

$$SB = \max_{1 \leq i \leq N} |S_i^*| \text{ and } WB = \max_{1 \leq i \leq N} |W_i^*|. \tag{2.4.1}$$

Note that the maximal statistic corresponds to the one with the smallest p-value, since the individual test statistics are identically distributed. These test statistics are suggested by the fact that H_0 in (2.2.3) is the intersection of the N subhypotheses $H_{0i} : a_i = 0$, $i = 1, ..., N$. The decision rule is then built from the logical equivalence that H_0 is false if any of its component subhypotheses is false; i.e., reject H_0 if any one of the separate tests, say $S_1^*, ..., S_N^*$, rejects it. Such tests are called *induced* tests of H_0; see Savin (1984). The results of Sidak (1967) show that no matter the covariance structure among the individual test statistics, the asymptotic marginal null distributions of SB and WB defined in (2.4.1) satisfy the inequalities

$$\Pr[SB \leq \omega_{\alpha^*/2}] \geq (1 - \alpha) \text{ and } \Pr[WB \leq \omega_{\alpha^*/2}] \geq (1 - \alpha), \tag{2.4.2}$$

where $\omega_{\alpha^*/2}$ is the upper $\alpha^*/2$ critical point of the standard normal distribution and $\alpha^* = 1 - (1-\alpha)^{1/N}$. This inequality gives a slight improvement over the Bonferroni inequality. The inequalities in (2.4.2) imply that the asymptotic *level* of the induced test of H_0 that compares either SB or WB to $\omega_{\alpha^*/2}$ is equal to α. So if the ordinary two-sided p-value of SB or WB is, say pv, then the multiplicity-adjusted two-sided p-value is calculated from the equation $pv^* = 1 - (1-pv)^N$. See Chow and Denning (1993) for an application of Sidak's results in a different context.

2.4.2 Sum-type tests

The sup-type tests based on the bounds in (2.4.2) allow for an arbitrary covariance structure in the cross-section of error terms. The cost of this extra flexibility is that those tests tend to be conservative and this behavior under the null effectively translates into relative power losses under the alternative hypothesis; see Section 4 for some simulation evidence.

Tests with the correct *size* that deliver more power can be obtained under the assumption that the error terms $(\varepsilon_{1t}, ..., \varepsilon_{Nt})$ are independent, conditional on $R_p = (r_{p1}, ..., r_{pT})'$. Note that the error terms need not be unconditionally independent. With that further assumption, the exact distributions of the sum-type statistics

$$SS = \sum_{i=1}^{N} S_i^{*2} \text{ and } WS = \sum_{i=1}^{N} W_i^{*2} \tag{2.4.3}$$

can be obtained (e.g. by numerical or simulation methods). A far more practical way to conduct inference is to note that since the S_i^*'s and W_i^*'s tend to be normal as m increases, the null distributions of SS and WS in (2.4.3) will tend to be chi-squared with N degrees of freedom. The results in Section 4 show that this approximation works extremely well.

2.4.3 Single-portfolio tests

As GRS remark, the hypothesis that $a_i = 0$, for all i, implies that all linear combinations of the a_i's equal zero. This means that if some linear combination of the N test assets (or portfolios of test assets) has a non-zero intercept, then the null hypothesis is false; i.e., portfolio p is not mean-variance efficient. Following that idea, consider aggregating the N equations that comprise the system in (2.2.2) to find the single-portfolio representation

$$\bar{r}_t = A + Br_{pt} + e_t, \quad t = 1, ..., T, \tag{2.4.4}$$

where $\bar{r}_t = \sum_{i=1}^N r_{it}$, $A = \sum_{i=1}^N a_i$, $B = \sum_{i=1}^N \beta_i$, and $e_t = \sum_{i=1}^N \varepsilon_{it}$. It is also possible to consider weighted sums. The weights, however, would need to be chosen on the basis of prior (non-sample) information to avoid introducing data-snooping size distortions; see Lo and MacKinlay (1990).

The construction of a test of H_0 in (2.2.3) based on A is reminiscent of the test in Bossaerts and Hillion (1995) based on $\hat{A} = \sum_{i=1}^{N} \hat{a}_i$. The same distribution-free approach as above can be used to detect whether the intercept A in (2.4.4) is different from zero under the following high-level assumptions:

Assumption 3. The density of $(e_1, ..., e_T)$ is symmetric around zero, conditional on $R_p = (r_{p1}, ..., r_{pT})'$. Further, $\Pr[r_{pt} = 0] = 0$ for $t = 1, ..., T$.

Assumption 4. The random variables e_t, $t = 1, ..., T$, are continuous, conditional on R_p.

These assumptions are the natural extensions of Assumptions 1 and 2 to the model in (2.4.4). As before, no restrictions are placed on the shape or the location of the distribution of r_{pt}. It is easy to see that if Assumption 1 holds and the errors are either cross-sectionally independent or at least exchangeable as Bossaerts and Hillion (1995) assume, then Assumption 3 is satisfied. A sufficient condition for Assumption 3 is that the joint cross-sectional density of error terms is reflective symmetric so that $(\varepsilon_{1t}, ..., \varepsilon_{Nt})$ has the same distribution as $(-\varepsilon_{1t}, ..., -\varepsilon_{Nt})$. In that case, we have that $\sum_{i=1}^{N} \varepsilon_{it}$ has the same distribution as $-\sum_{i=1}^{N} \varepsilon_{it}$; i.e., e_t is symmetrically distributed around zero. So although not completely unrestricted, reflective symmetry allows some forms of covariation in the cross-section of error terms.

Our distribution-free approach yields the following versions of the sign and Wilcoxon test statistics:

$$S = \sum_{t=1}^{m} s[Z_t] \text{ and } W = \sum_{t=1}^{m} s[Z_t] Rank(|Z_t|), \tag{2.4.5}$$

where $Rank(|Z_t|)$ is the rank of $|Z_t|$ when $|Z_1|, ..., |Z_m|$ are placed in ascending order of magnitude. Here Z_t is defined as

$$Z_t = \left(\frac{\bar{r}_{t+m}}{r_{p,t+m}} - \frac{\bar{r}_t}{r_{pt}} \right) \left(\frac{r_{pt} - r_{p,t+m}}{r_{pt} r_{p,t+m}} \right),$$

for $t = 1, ..., m$. Under the null hypothesis, the sign and Wilcoxon statistics in (2.4.5) based on Z_t follow exactly the binomial distribution (with parameters m and 1/2) and the distribution of the Wilcoxon variate, respectively. As before, the standardized versions

$$S^* = \frac{S - m/2}{\sqrt{m/4}} \text{ and } W^* = \frac{W - m(m+1)/4}{\sqrt{m(m+1)(2m+1)/24}} \tag{2.4.6}$$

have limiting standard normal distributions under H_0. Of course, the signed rank tests built on Z_t have no power to detect mean-variance inefficiencies of portfolio p in directions away from the null where $\sum_{i=1}^{N} a_i = 0$. This possibility, however, can easily be checked by examining the estimates $\hat{a}_1, ..., \hat{a}_N$; see Section 5 for an empirical illustration with size portfolios.

The possibility of losing power depending on whether the (weighted) a_i's tend to cancel out is a concern for any approach that uses linear combinations of assets. For example, parametric tests of the CAPM are usually applied to portfolio groupings of stocks in order to have N much less than T. As Shanken (1996) notes, this has the potential effect of reducing the residual variances and increasing the precision with which the a_i's are estimated. On the other hand, as Roll (1979) points out, individual stock expected return deviations under the alternative can cancel out in portfolios, which would reduce power. So the expected power of tests to detect mean-variance inefficiencies necessarily depends on how the portfolios are formed; see Sentana (2008) for more on the effects of portfolio composition on test power.

2.5 Simulation results

This section reports the results of some simulation experiments to contrast the behavior of the proposed distribution-free tests with several standard test procedures. The first of these benchmarks is the GRS test in (2.2.4). The second one, denoted J_2^{MC}, is an MC test based on the LR statistic in (2.2.6), following the BDK approach. The other benchmarks for comparison purposes are the usual LR test, an adjusted LR test, and a test based on the Generalized Method of Moments (GMM). The latter is a particularly important benchmark here, since in principle it is "robust" to non-normality and heteroskedasticity of returns.

The LR test, J_2, is applied following the usual practice which is to base it on asymptotic theory. The large-sample result is that the null distribution of J_2 tends to that of a chi-square variate with N degrees of freedom, χ^2_N. Jobson and Korkie (1982) suggest an adjustment to J_2 in order to improve its finite-sample size properties when compared against the χ^2_N distribution. The adjusted statistic is

$$J_3 = \frac{T - (N/2) - 2}{T} J_2, \tag{2.5.1}$$

which, under H_0, also follows an asymptotic χ^2_N distribution.

MacKinlay and Richardson (1991) develop tests of mean-variance efficiency using a GMM framework. Define r_t as the $N \times 1$ vector of excess returns for the N test assets observed at time t. The moments of interest are those of the $2N \times 1$ vector

$$f_t(\theta) = h_t \otimes \varepsilon_t(\theta), \tag{2.5.2}$$

where $h_t = (1, r_{pt})'$ and $\varepsilon_t(\theta) = r_t - a - \beta r_{pt}$. Here $\theta = (a', \beta')'$, $a = (a_1, ..., a_N)'$, and $\beta = (\beta_1, ..., \beta_N)'$. The symbol $\otimes$ refers to the Kronecker product. The model specification in (2.2.2) implies the moment conditions $E(f_t(\theta_0)) = 0$, where θ_0 is the true parameter vector. The system in (2.5.2) is exactly identified which implies that the GMM procedure yields the same estimates of θ as does OLS applied equation by equation, which in turn are equivalent to the maximum likelihood estimates in (2.2.2). The GMM-based Wald test statistic is

$$J_4 = T\hat{a}' \left[R \left(\hat{D}'\hat{S}^{-1}\hat{D} \right)^{-1} R' \right]^{-1} \hat{a}, \tag{2.5.3}$$

where $R = (1, 0) \otimes I_N$, with I_N as the $N \times N$ identity matrix. Here $T^{-1}(\hat{D}'\hat{S}^{-1}\hat{D})^{-1}$ is a consistent estimator of the covariance matrix of the GMM estimator $\hat{\theta}$. The component matrix $\hat{D}$ is computed as

$$\hat{D} = - \begin{bmatrix} 1 & \bar{r}_p \\ \bar{r}_p & (\hat{\sigma}_p^2 + \bar{r}_p^2) \end{bmatrix} \otimes I_N$$

and $\hat{S}$, which is a consistent estimator of $\sum_{\ell=-\infty}^{+\infty} E[f_t(\theta_0) f_{t-\ell}(\theta_0)']$, can be computed with the Newey and West (1987) procedure. Under the null hypothesis H_0, the statistic J_4 follows an asymptotic χ^2_N distribution. Following MacKinlay and Richardson (1991), the J_4 statistic in (2.5.3) is scaled by $(T - N - 1)/T$, which was suggested to improve the finite-sample null behavior of the GMM Wald test.

We also consider the parametric J tests applied to the single-portfolio representation in (2.4.4). Those tests are denoted J_1^*, J_2^{*MC}, J_2^*, J_3^*, and J_4^* and they will allow us to see the relative power gains that result from the non-parametric approach versus those that the linear combination induces. Note that since $\log(1 + x) \approx x$ for x small, we expect J_3^* to mimic J_1^*.

All the tests compared here are invariant with respect to the β_i's, so there is no need to specify their values. The MC tests are applied assuming the distribution of the error terms is completely known and the p-values are based on 100 MC draws. We consider sample sizes $T = 30, 60, 120$, and 240. For the number of test assets, we consider values of $N = 10, 20$, and 30. Note that J_1, J_2^{MC}, J_2, J_3, and J_4 cannot be computed when $N \geq T$. (Those cases are abbreviated n.a. in the tables.) The GMM Wald test, J_4, cannot even be computed with $T = 60$ and $N = 30$. The reason is that the matrix $\hat{S}$ is singular whenever $2N = T$. All the other tests do not have these limitations on N relative to T. In order to have a sharp contrast between the power functions of each test, the parameters a_i are set to 0.07 under the alternative hypothesis. The parametric MC tests are exact here, but there is no such theoretical guarantee with all the other parametric tests across error specifications. So to ensure meaningful comparisons, the power results for the non-randomized parametric tests are based on size-corrected critical values. Finally, each experiment comprises 1000 replications of each return generating process.

Two specifications of model (2.2.2) are considered. The first example, though completely unrealistic, is constructed to ensure the exactness of the GRS Wald test, J_1. The error terms ε_{it} are drawn randomly from the standard normal distribution and so are portfolio-p's excess returns, r_{pt}, independently of the errors. MacKinlay and Richardson (1991) use a similar design. We also tried introducing coskewness by drawing r_{pt} from a χ_1^2 distribution. The results were essentially the same as in the symmetric case.

Table 1 reports the empirical rejection rates for this homoskedastic example. As expected, the parametric J_1 and J_2^{MC} tests, the distribution-free sum-type SS and WS tests, and the single-portfolio J_1^*, J_2^{*MC}, S^*, and W^* tests have empirical sizes close to the nominal 5% level. Note how well the use of critical values from the chi-squared distribution works for the SS and WS tests, and those of the normal distribution for the S^* and W^* tests. In general, all the single-portfolio J^* tests also have rejection rates close to the nominal level. We see though that even in this i.i.d. case, the LR J_2 test suffers serious size distortions when T is relatively small and N

increases. For example, when N increases from 10 to 20 with $T = 30$, the empirical size of the J_2 test increases from about 15% to 60%—twelve times the nominal level.

The bottom portion of Table 1 shows the relative power results. Note that J_1, J_2, and J_3 have identical size-corrected powers, since they are all related via monotonic transformations. The same is true of their single-portfolio counterparts. As expected, the J_2^{MC} and J_2^{*MC} tests behave like the J_1 and J_1^* tests, respectively. It is immediately clear that SB and WB are not as powerful as the other tests. They nevertheless have power as T increases. The most striking result though comes from comparing the S^* and W^* tests with the parametric tests. For example, the W^* test has power that is either comparable or even greater in some instances than all the parametric J tests. The power advantage that the linear combination induces in this setup is apparent when comparing any test to its single-portfolio counterpart. The power performance of the S^* and W^* tests is quite remarkable considering their non-parametric nature and the fact that only half the sample is effectively used to detect departures from the null hypothesis.

The second specification of model (2.2.2) is a more realistic description of asset returns. Following MacKinlay and Richardson (1991), we consider a case of contemporaneous conditional heteroskedasticity in which the variance of ε_{it} depends on the value of r_{pt}. In their example, the error terms are essentially governed by an equation of the form $\varepsilon_{it} = \eta_{it}\sqrt{(r_{pt} - \mu_p)^2/\sigma_p^2}$, where η_{it} is the innovation term. As in standard GARCH models, this form does not allow an asymmetric response to shocks in r_{pt}; i.e., the conditional variance of ε_{it} is the same whether $(r_{pt} - \mu_p)$ is positive or negative. Here we consider a specification in the spirit of Nelson's (1991) exponential GARCH model that allows for asymmetric responses to shocks. The innovations η_{it} are standard normal and we let $\varepsilon_{it} = \eta_{it}\sqrt{\exp(\lambda_0 + \lambda_1 r_{pt})}$. It should be noted that such a specification finds empirical support in Duffee (1995, 2001). The returns r_{pt} are as in the first example and we set $\lambda_0 = 0$, $\lambda_1 = 2$. It is easy to see that this specification generates ε_{it}'s with excess kurtosis.

Table 2 reports the results for the heteroskedastic example. From the top portion where $a_i = 0$, we see that all the non-randomized parametric J tests suffer serious size distortions, and that these worsen as N increases for any given T. When $N = 10$, $T = 60$ the parametric J tests all have empirical sizes in excess of 20%. The probability of a Type I error with J_1, J_2, and J_3 exceeds 60% when N is increased to 30. The severity of the size distortions for the GMM-based J_4 test is also quite surprising since it accomodates conditional heteroskedasticity, at least in principle. In sharp contrast, there is a closer agreement between the nominal and empirical rejection probabilities for the single-portfolio J^* tests. The MC tests do not have an over-rejection problem here, since they are applied assuming the form of heteroskedasticity is known. The distribution-free tests also behave as expected with rejection rates close to the nominal 5% level, but unlike the MC tests this is achieved without assuming anything about the heterogeneity in variances.

From the bottom portion of Table 2, we see that the distribution-free tests have very good power when compared to the parametric tests. The sum-type SS and WS tests have power which is comparable to that of the J tests, including the MC tests. It should be emphasized that the size-corrected tests are not feasible tests in practice. They are merely used here as theoretical benchmarks for the new tests. Comparing the single-portfolio tests, we see again that the S^* and W^* tests in this heteroskedastic case tend to be more powerful than the parametric tests, often by a sizeable margin. This result is in line with the discussion in Randles and Wolfe (1979, Ch. 4) about the power of the sign and Wilcoxon signed rank tests relative to the Student-t test in a location model under a variety of distributions. Here we have another example that illustrates the conventional wisdom that non-parametric tests tend to perform well in the presence of excess kurtosis—a well-known feature of asset returns.

2.6 Empirical illustration

The new distribution-free inference procedures are illustrated by tests of the mean-variance efficiency of a stock market index: the value-weighted index from the University of Chicago's Center for Research in Security Prices (CRSP). The data consist of monthly returns on ten portfolios for the period covering January 1965 to December 2006. Stocks listed on the New York Stock Exchange and on the American Stock

Exchange are allocated to 10 portfolios based on the market value of equity. The one-month US Treasury bill return is used as the riskless rate of return. These data definitions are the same as in Campbell, Lo, and MacKinlay (1997, Ch. 5), except that we extend the sample end point.

Table 3 shows the empirical test results for the entire forty-two-year period, seven five-year subperiods and a seven-year subperiod, and three ten-year subperiods and a twelve-year subperiod. It is quite common in this literature to test the CAPM over subperiods out of concerns about parameter stability. Here the choice of subperiods follows that in Campbell, Lo, and MacKinlay. Columns 2–6 report the results of the parametric J tests; columns 7 and 8 report the results of the sup-type SB and WB tests; columns 9 and 10 report the results of the sum-type SS and WS tests; columns 11–15 report the results of the single-portfolio J^* tests; and columns 16 and 17 report the results of the single-portfolio S^* and W^* tests. Here the MMC test procedure of BDK is applied assuming Student-t errors with unknown degrees of freedom. In each case, we set $\alpha_1 = 0.025$ so the level of the first-step confidence interval for the degrees-of-freedom parameter is 97.5%. These confidence intervals are reported in square brackets. The reported MMC p-values should then be referred to a 2.5% cutoff if one has in mind an overall 5%-level test.

The parametric J test results tend to reject the Sharpe-Lintner CAPM, with p-values less than 6% for the entire sample period. The J^* tests reject the null hypothesis with even smaller p-values in this case. The exceptions among the parametric J tests are the MMC tests which tend to favor a non-rejection. Note that the confidence intervals associated with the MMC tests are tight at quite low values for the degrees-of-freedom parameter. This suggests the presence of high kurtosis and shows that normality is definitely rejected. In contrast, the p-values for the distribution-free tests provide very clear evidence in favor of mean-variance efficiency.

The parametric J tests and distribution-free tests results align more closely over the five-year subperiods with both sets of tests tending to reject the null in the first three five-year subperiods, then tending to not reject over the next four five-year subperiods, and finally tending to reject again the null in the last seven-year subperiod. The single-portfolio J^* tests tend to be in agreement with the other tests in every one of those subperiods except in the subperiod 1/70–12/74, where the J^* tests indicate a clear non-rejection but all the other tests seem to favor a rejection. The results for the ten-year subperiods, reported in the bottom portion of Table 3, are more in line with those for the entire sample period. In the first three of those subperiods, the parametric tests tend to reject the null, while the distribution-free tests do not reject the null hypothesis of efficiency in two out of those four subperiods. Here the single-portfolio tests agree on rejections in the second and fourth subperiods, and non-rejections in the first and third subperiods.

Besides the obvious sensitivity to the choice of sample period, a comparison of the results across test statistics reveals that parametric versus non-parametric inference can differ markedly. Indeed, in many cases where the distribution-free tests indicate a clear non-rejection with large p-values, the parametric tests indicate the diametrical opposite with quite small p-values. The full sample results are a case in point.

Table 4 shows the OLS estimates of the intercept terms for each equation comprising the system in (2.2.2) and the equation in (2.4.4). The columns labeled P_1 to P_{10} show the estimates of a_i corresponding to the ten portfolios and the last column, labeled P_A, shows the estimates of the aggregate portfolio intercept $A = \sum_{i=1}^{10} a_i$. All the reported intercept values are multiplied by 1000. It is interesting to note that the individual intercept terms for portfolios P_1 to P_9 tend to all have the same sign, and the estimated intercept with the opposing sign for portfolio P_{10}—the largest cap-based decile portfolio—is quite small in magnitude compared to $\sum_{i=1}^{9} \hat{a}_i$. These results are evidence that the non-rejections by the distribution-free single-portfolio S^* and W^* tests seen in Table 3 are not due to the a_i's canceling out.

Table 5 reports the covariance matrix for the market model residuals, where the entries are multiplied by 1000. This matrix is an estimate of the covariances across portfolio returns once the effects of the market portfolio have been removed. The low covariance values suggest that assuming conditional independence in the cross-section of returns may be fairly innocuous in this application. Indeed, recall that the sum-type SS and WS tests assume that the error terms $(\varepsilon_{1t}, ..., \varepsilon_{Nt})$ are independent, conditional on R_p. The fact that those test statistics are not significant when computed with the full sample constitutes an implicit non-rejection of the conditional independence assumption. As a further check, we computed the Breusch and Pagan (1980) Lagrange multiplier test for the diagonality of the error covariance matrix of a seemingly unrelated equations system. The computed p-value of that test is 0.99, which clearly shows the joint statistical insignificance of the off-diagonal entires in Table 5. These results support not only the validity of the sum-type SS and WS tests but also that of the single-portfolio S^* and W^* tests. In light of the simulation results, the most natural interpretation of the empirical findings is that the CRSP value-weighted index appears consistent with the mean-variance efficiency hypothesis over the full sample period.

2.7 Conclusion

The mean-variance efficiency of a given portfolio is typically assessed with tests based on OLS or GMM. The reliability of such test procedures depends on several parametric assumptions about the model's error terms, such as the finiteness and stationarity of moments. Exact parametric tests rely on an even stronger assumption about the actual distribution of the error terms. For instance, the well-known GRS Wald test assumes that the errors are multivariate Gaussian.

In this paper, we have proposed new non-randomized tests of mean-variance efficiency whose exactness does not depend on any parametric assumptions. The finite-sample validity of the new tests rests on the assumption that the joint temporal error density is symmetric around zero. This leaves open the possibility of conditional heteroskedasticity and even outliers. In fact, no restrictions are placed on the degree of non-normality or the degree of heterogeneity and dependence of the conditional variances. Not even the existence of moments is required.

The proposed tests were evaluated against commonly used parametric tests in some simulation experiments. The numerical results reveal that the parametric procedures suffer serious size distortions in the presence of contemporaneous conditional heteroskedasticity, especially when the number of test assets is large and/or the time history is relatively short. The MC approach of BDK can avoid this problem, but only if the precise form of heterogeneity is known. Since it has long been recognized that asset returns depart from homogeneous conditions, the new tests of the mean-variance efficiency hypothesis developed here offer a valid and useful distribution-free testing alternative to potentially very misleading parametric procedures.

Appendix

Proof of Proposition 1. Under Assumption 1, the elements of R_p are all different from zero with probability one, so the quantities ε_{it}/r_{pt}, $t = 1, ..., T$, are well defined. Given R_p, we also have under Assumption 1 that

$$(\varepsilon_{i1}/r_{p1}, \varepsilon_{i2}/r_{p2}, ..., \varepsilon_{iT}/r_{pT}) \overset{d}{=} (-\varepsilon_{i1}/r_{p1}, \varepsilon_{i2}/r_{p2}, ..., \varepsilon_{iT}/r_{pT}) \overset{d}{=} ...$$

$$\overset{d}{=} (-\varepsilon_{i1}/r_{p1}, -\varepsilon_{i2}/r_{p2}, ..., -\varepsilon_{iT}/r_{pT}), \tag{2.7.1}$$

where all 2^T such terms appear in this string of equalities in distribution. In particular, we have

$$(\varepsilon_{i1}/r_{p1}, \varepsilon_{i2}/r_{p2}, ..., \varepsilon_{iT}/r_{pT}) \overset{d}{=} (-\varepsilon_{i1}/r_{p1}, -\varepsilon_{i2}/r_{p2}, ..., -\varepsilon_{iT}/r_{pT}).$$

Define $\delta[\eta_1, \eta_2, ..., \eta_T] = \big((\eta_{m+1} - \eta_1), (\eta_{m+2} - \eta_2), ..., (\eta_T - \eta_m)\big)$. It follows that

$$\delta\left[\varepsilon_{i1}/r_{p1}, \varepsilon_{i2}/r_{p2}, ..., \varepsilon_{iT}/r_{pT}\right] \overset{d}{=} \delta\left[-\varepsilon_{i1}/r_{p1}, -\varepsilon_{i2}/r_{p2}, ..., -\varepsilon_{iT}/r_{pT}\right]$$

or

$$\big((\varepsilon_{i,m+1}/r_{p,m+1} - \varepsilon_{i1}/r_{p1}), ..., (\varepsilon_{iT}/r_{pT} - \varepsilon_{im}/r_{pm})\big) \overset{d}{=}$$

$$\big(- (\varepsilon_{i,m+1}/r_{p,m+1} - \varepsilon_{i1}/r_{p1}), ..., -(\varepsilon_{iT}/r_{pT} - \varepsilon_{im}/r_{pm})\big),$$

since $X \overset{d}{=} Y$ implies $U[X] \overset{d}{=} U[Y]$ for any measurable function $U[\cdot]$ defined on the common support of X and Y (Randles and Wolfe 1979, Theorem 1.3.7). This last result is used here conditional on R_p so that

$$\Big((\varepsilon_{i,m+1}/r_{p,m+1} - \varepsilon_{i1}/r_{p1})x_{p1}, ..., (\varepsilon_{iT}/r_{pT} - \varepsilon_{im}/r_{pm})x_{pm}\Big) \stackrel{d}{=}$$

$$\Big(- (\varepsilon_{i,m+1}/r_{p,m+1} - \varepsilon_{i1}/r_{p1})x_{p1}, ..., -(\varepsilon_{iT}/r_{pT} - \varepsilon_{im}/r_{pm})x_{pm}\Big),$$

since the x_{pt}'s defined in (3.4.4) are just functions of $(r_{p1}, ..., r_{pT})'$. In turn,

$$E\Big[s\big[\,(\varepsilon_{i,t+m}/r_{p,t+m} - \varepsilon_{it}/r_{pt})\,x_{pt}\big]\Big] = E\Big[s\big[-(\varepsilon_{i,t+m}/r_{p,t+m} - \varepsilon_{it}/r_{pt})\,x_{pt}\big]\Big]$$

or

$$\Pr\big[\,(\varepsilon_{i,t+m}/r_{p,t+m} - \varepsilon_{it}/r_{pt})\,x_{pt} > 0\big] = \Pr\big[\,(\varepsilon_{i,t+m}/r_{p,t+m} - \varepsilon_{it}/r_{pt})\,x_{pt} < 0\big] = 1/2, \tag{2.7.2}$$

for $t = 1, 2, ..., m$, since $\Pr\left[(\varepsilon_{i,t+m}/r_{p,t+m} - \varepsilon_{it}/r_{pt})\,x_{pt} = 0\right] = 0$ under Assumption 2. Note that (2.7.2) holds conditional on R_p as well as unconditionally. $\square$

Proof of Proposition 2. By applying the function $\delta[\cdot]$ defined in the proof of Proposition 1 to the string of equalities in distribution in (2.7.1), we see that

$$\Big((\varepsilon_{i,m+1}/r_{p,m+1} - \varepsilon_{i1}/r_{p1}), (\varepsilon_{i,m+2}/r_{p,m+2} - \varepsilon_{i2}/r_{p2}), ..., (\varepsilon_{iT}/r_{pT} - \varepsilon_{im}/r_{pm})\Big) \stackrel{d}{=}$$

$$\Big(-(\varepsilon_{i,m+1}/r_{p,m+1} - \varepsilon_{i1}/r_{p1}), (\varepsilon_{i,m+2}/r_{p,m+2} - \varepsilon_{i2}/r_{p2}), ..., (\varepsilon_{iT}/r_{pT} - \varepsilon_{im}/r_{pm})\Big) \stackrel{d}{=}$$

$$\Big((\varepsilon_{i,m+1}/r_{p,m+1} - \varepsilon_{i1}/r_{p1}), -(\varepsilon_{i,m+2}/r_{p,m+2} - \varepsilon_{i2}/r_{p2}), ..., (\varepsilon_{iT}/r_{pT} - \varepsilon_{im}/r_{pm})\Big) \stackrel{d}{=}$$

$$\Big(-(\varepsilon_{i,m+1}/r_{p,m+1} - \varepsilon_{i1}/r_{p1}), -(\varepsilon_{i,m+2}/r_{p,m+2} - \varepsilon_{i2}/r_{p2}), ..., (\varepsilon_{iT}/r_{pT} - \varepsilon_{im}/r_{pm})\Big) \stackrel{d}{=}$$

$$\vdots$$

$$\stackrel{d}{=} \Big(-(\varepsilon_{i,m+1}/r_{p,m+1} - \varepsilon_{i1}/r_{p1}), -(\varepsilon_{i,m+2}/r_{p,m+2} - \varepsilon_{i2}/r_{p2}), ..., -(\varepsilon_{iT}/r_{pT} - \varepsilon_{im}/r_{pm})\Big),$$

where all 2^m such terms appear in this string of equalities in distribution. Let $E = \Big((\varepsilon_{i,m+1}/r_{p,m+1} - \varepsilon_{i1}/r_{p1})x_{p1}, ..., (\varepsilon_{iT}/r_{pT} - \varepsilon_{im}/r_{pm})x_{pm}\Big)$. It follows that the 2^m

different values that the vector $s[E] = \Big(s\big[(\varepsilon_{i,m+1}/r_{p,m+1} - \varepsilon_{i1}/r_{p1})x_{p1}\big], ..., s\big[(\varepsilon_{iT}/r_{pT} - \varepsilon_{im}/r_{pm})x_{pm}\big]\Big)$ may take in $\{0,1\}^m$ have the same probability $(1/2)^m$. Therefore, the elements of $s[E]$ are mutually independent. Note that this result holds conditional on R_p as well as unconditionally. Define d_j to be the position of the integer j in the realization of the vector $\big(Rank(|z_{i1}|), ..., Rank(|z_{im}|)\big)$, $j = 1, ..., m$, so that

$$\sum_{t=1}^{m} s[z_{it}]\varphi\big[Rank(|z_{it}|)\big] = \sum_{j=1}^{m} s[z_{i,d_j}]\varphi[j].$$

Conditional on $|E| = \Big(|(\varepsilon_{i,m+1}/r_{p,m+1} - \varepsilon_{i1}/r_{p1})x_{p1}|, ..., |(\varepsilon_{iT}/r_{pT} - \varepsilon_{im}/r_{pm})x_{pm}|\Big)$, the vector of scores is a fixed permutation of $\big(\varphi[1], ..., \varphi[m]\big)$. So under the null hypothesis and conditional on $|E|$, it follows that

$$\sum_{j=1}^{m} s[z_{i,d_j}]\varphi[j] \overset{d}{=} \sum_{j=1}^{m} B_j\varphi[j],$$

since $\big(s[z_{i,d_1}], ..., s[z_{i,d_m}]\big) \overset{d}{=} (B_1, B_2, ..., B_m)$, where $B_1, ..., B_m$ are mutually independent uniform Bernoulli variables on $\{0,1\}$. Moreover, given the symmetry established in Proposition 1, we have under the null that $s\left[z_t\right]$ is independent of R_t^+ and thus of $\varphi\left[R_t^+\right]$ (Randles and Wolfe 1979, Lemma 2.4.2). Therefore, under the null, it is the case also unconditionally that

$$SR_i = \sum_{t=1}^{m} s[z_{it}]\varphi\big[Rank(|z_{it}|)\big] \overset{d}{=} \sum_{j=1}^{m} B_j\varphi[j],$$

since the distribution of $\sum_{j=1}^{m} B_j\varphi[j]$ does depend on $|E|$. $\square$

References

Beaulieu, M.-C., J.-M. Dufour, and L. Khalaf. 2007. "Multivariate tests of mean-variance efficiency with possibly non-Gaussian errors: an exact simulation-based approach." *Journal of Business and Economic Statistics* 25: 398–410.

Berk, J. 1997. "Necessary conditions for the CAPM." *Journal of Economic Theory* 73: 245–257.

Blattberg, R. and N. Gonedes. 1974. "A comparison of the stable and student distributions as statistical models of stock prices." *Journal of Business* 47: 244–280.

Bollerslev, T. 1986. "Generalized autoregressive conditional heteroskedasticity." *Journal of Econometrics* 31: 307–327.

Bossaerts, P. and P. Hillion. 1995. "Testing the mean-variance efficiency of well-diversified portfolios in very large cross-sections." *Annales d'Économie et de Statistique* 40: 93–124.

Breusch, T.S. and A.R. Pagan. 1980. "The Lagrange multiplier test and its applications to model specification in econometrics." *Review of Economic Studies* 47: 239–253.

Campbell, J.Y., A.W. Lo, and A.C. MacKinlay. 1997. The Econometrics of Financial Markets. Princeton University Press, New Jersey.

Chamberlain, G. 1983. "A characterization of the distributions that imply mean-variance utility functions." *Journal of Economic Theory* 29: 185–201.

Chow, K.V. and K.C. Denning. 1993. "A simple multiple variance ratio test." *Journal of Econometrics* 58: 385–401.

Duffee, G.R. 1995. "Stock returns and volatility: A firm-level analysis." *Journal of Financial Economics* 37: 399–420.

Duffee, G.R. 2001. "Asymmetric cross-sectional dispersion in stock returns: Evidence and implications." Federal Reserve Bank of San Francisco Working Paper No. 2000-18.

Dufour, J.-M. and L. Khalaf. 2001. Monte Carlo test methods in econometrics. In: Baltagi, B. (Ed.), Companion to Theoretical Econometrics. Basil Blackwell, Oxford.

Dufour, J.-M. and L. Khalaf. 2002. "Simulation based finite and large sample tests in multivariate regressions." *Journal of Econometrics* 111: 303–322.

Fama, E. 1965. "The behavior of stock market prices." *Journal of Business* 38: 34–105.

Gibbons, M.R., S.A. Ross, and J. Shanken. 1989. "A test of the efficiency of a given portfolio." *Econometrica* 57: 1121–1152.

Hollander, M. and D.A. Wolfe. 1973. Nonparametric Statistical Methods. Wiley, New York.

Hsu, D.A. 1982. "A Bayesian robust detection of shift in the risk structure of stock market returns." *Journal of the American Statistical Association* 77: 29–39.

Ingersoll, J. 1987. *Theory of Financial Decision Making.* Rowman & Littlefield.

Jobson, J.D. and B. Korkie. 1982. "Potential Performance and Tests of Portfolio Efficiency." *Journal of Financial Economics* 10: 433–466.

Lintner, J. 1965. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." *Review of Economics and Statistics* 47: 13–37.

Lo, A.W. and A.C. MacKinlay. 1990. "Data-snooping biases in tests of financial asset pricing models." *Review of Financial Studies* 3: 431–467.

Luger, R. 2003. "Exact non-parametric tests for a random walk with unknown drift under conditional heteroscedasticity." *Journal of Econometrics* 115: 259–276.

MacKinlay, A.C. and M.P. Richardson. 1991. "Using Generalized Method of Moments to test mean-variance efficiency." *Journal of Finance* 46: 511–527.

Nelson, D.B. 1991. "Conditional heteroskedasticity in asset returns: a new approach." *Econometrica* 59: 347–370.

Newey, W. and K. West. 1987. "A simple, positive semi-definite, heteroskedasticity and autocorrelation covariance matrix." *Econometrica* 55: 703–708.

Owen, J. and R. Rabinovitch. 1983. "On the class of elliptical distributions and their applications to the theory of portfolio choice." *Journal of Finance* 38: 745–752.

Randles, R.H. and D.A. Wolfe. 1979. Introduction to the Theory of Nonparametric Statistics. Wiley, New York.

Roll, R. 1979. "A Reply to Mayers and Rice (1979)." *Journal of Financial Economics* 7: 391–400.

Savin, N.E. 1984. "Multiple hypothesis testing." In Griliches, Z. and M.D. Intrilligator (Eds.), *Handbook of Econometrics* Vol. 2, North-Holland, Amsterdam, 827–879.

Sentana, E. 2008. "The econometrics of mean-variance efficiency tests: a survey." CEMFI Working Paper No. 0807.

Shanken, J. 1996. "Statistical Methods in Tests of Portfolio Efficiency: A Synthesis." In *Handbook of Statistics, Vol. 14: Statistical Methods in Finance,* edited by G.S. Maddala and C.R. Rao. Amsterdam: North-Holland.

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

Sidak, Z. 1967. "Rectangular confidence regions for the means of multivariate normal distributions." *Journal of the American Statistical Association* 62: 626–633.

Theil, H. 1971. Principles of Econometrics. Wiley, New York.

Table 2.1: **Empirical size and power comparisons in the seemingly unrelated equations model with homoskedastic errors.**

N	T	J_1	J_2^{MC}	J_2	J_3	J_4	SB	WB	SS	WS	J_1^*	J_2^{*MC}	J_2^*	J_3^*	J_4^*	S^*	W^*
Under $H_0 : a_i = 0,\ i = 1, ..., N$																	
10	30	5.0	4.7	16.5	5.0	13.8	6.6	2.6	3.9	4.0	4.8	5.5	6.4	4.7	6.9	2.5	4.2
	60	4.0	4.8	7.4	4.0	6.4	5.5	3.5	4.8	4.4	4.0	4.0	4.3	4.0	4.4	5.1	4.9
	120	4.9	4.9	7.3	4.9	6.5	6.3	5.5	5.3	5.1	3.9	4.0	4.1	3.9	4.0	4.5	5.0
	240	3.8	3.7	4.6	3.8	4.6	3.0	4.5	4.1	4.6	3.9	3.7	3.9	3.9	3.9	4.5	4.2
20	30	4.7	4.9	63.1	12.4	n.a.	2.1	1.8	3.3	3.5	4.7	4.3	5.9	4.7	6.1	4.2	4.9
	60	4.7	5.3	22.8	5.3	12.7	2.8	3.1	4.6	4.7	4.5	4.2	5.1	4.5	5.2	4.1	4.6
	120	5.1	5.1	11.5	5.2	8.8	5.1	3.7	4.2	4.6	6.0	6.0	6.3	6.0	6.3	4.7	4.5
	240	5.3	4.9	6.6	5.3	5.9	5.0	4.5	4.8	4.8	4.7	5.4	4.9	4.7	4.8	5.3	5.2
30	30	n.a.	n.a.	n.a.	n.a.	n.a.	2.8	1.2	4.6	3.1	4.4	4.5	5.7	4.4	5.7	3.5	4.1
	60	4.4	4.2	45.9	6.6	n.a.	3.1	3.0	4.7	5.0	4.8	4.7	5.5	4.8	5.6	3.8	5.4
	120	4.8	4.8	16.3	5.4	8.8	2.8	3.2	4.6	4.7	4.3	5.0	4.4	4.3	4.3	5.2	4.7
	240	4.5	4.5	8.5	4.5	6.5	3.6	4.9	5.3	5.0	5.1	5.6	5.3	5.1	5.2	5.5	5.1
Under $H_1 : a_i = 0.07,\ i = 1, ..., N$																	
10	30	7.0	7.2	7.0	7.0	7.9	6.8	2.8	4.2	5.6	20.6	20.3	20.6	20.6	20.4	6.9	10.4
	60	16.3	13.7	16.3	16.3	16.7	6.6	5.1	7.6	9.3	40.5	34.7	40.5	40.5	40.0	11.2	13.7
	120	29.0	26.6	29.0	29.0	28.3	9.0	8.8	8.9	10.2	71.3	64.5	71.3	71.3	70.9	23.2	26.6
	240	63.2	58.5	63.2	63.2	63.4	12.0	13.4	16.8	20.4	93.6	92.2	93.6	93.6	93.6	45.7	49.6
20	30	7.3	6.9	7.3	7.3	n.a.	2.1	2.2	5.4	6.3	38.8	36.2	38.8	38.8	38.5	12.0	16.0
	60	16.2	15.1	16.2	16.2	16.3	3.7	4.5	7.4	8.0	66.0	63.5	66.0	66.0	65.8	21.0	27.0
	120	43.2	41.3	43.2	43.2	43.4	8.8	9.5	11.5	14.0	91.6	92.4	91.6	91.6	91.7	43.5	49.0
	240	83.1	83.2	83.1	83.1	82.8	12.7	13.6	25.6	28.5	99.9	99.9	99.9	99.9	99.9	70.0	75.5
30	30	n.a.	n.a.	n.a.	n.a.	n.a.	2.8	1.5	6.1	7.2	51.2	47.4	51.2	51.2	50.5	13.7	17.4
	60	19.5	18.3	19.5	19.5	n.a.	5.3	3.4	6.8	8.8	83.2	82.8	83.2	83.2	82.8	27.6	34.5
	120	50.0	48.0	50.0	50.0	52.0	6.0	8.9	15.0	17.8	99.1	98.8	99.1	99.1	99.1	56.9	63.3
	240	92.1	90.7	92.1	92.1	92.2	9.8	12.4	28.7	34.5	100.0	100.0	100.0	100.0	100.0	85.1	90.8

Note: Nominal level is 0.05 and entries are percentage rates. Results based on 1000 replications. The abbreviation n.a. stands for not applicable.

Table 2.2: **Empirical size and power comparisons in the seemingly unrelated equations model with contemporaneous heteroskedastic errors.**

N	T	J_1	J_2^{MC}	J_2	J_3	J_4	SB	WB	SS	WS	J_1^*	J_2^{*MC}	J_2^*	J_3^*	J_4^*	S^*	W^*
Under $H_0 : a_i = 0,\ i = 1, ..., N$																	
10	30	25.9	4.6	50.7	27.7	41.6	6.0	2.1	3.8	4.3	8.1	4.7	9.6	8.1	7.4	2.9	3.7
	60	23.0	5.0	31.4	23.2	25.4	3.8	3.4	3.5	4.2	6.6	4.1	7.4	6.6	6.4	2.7	4.2
	120	15.5	5.8	19.5	15.5	15.5	5.8	4.2	3.7	4.3	6.1	5.4	6.5	6.1	5.4	5.6	4.4
	240	9.6	4.9	10.9	9.6	9.0	3.2	3.7	4.4	4.2	4.4	5.2	4.5	4.4	4.6	4.5	4.3
20	30	45.2	4.1	96.1	62.9	n.a.	2.0	1.8	4.6	4.9	7.0	4.9	8.3	7.0	6.8	4.1	5.4
	60	45.1	4.8	66.8	46.7	55.3	3.5	3.2	4.7	5.0	6.5	4.5	7.1	6.5	5.2	2.8	4.4
	120	29.9	4.5	41.0	30.2	33.1	5.5	4.5	5.6	5.4	6.7	5.0	6.8	6.7	5.5	5.4	4.1
	240	16.9	4.8	20.3	16.9	17.1	4.7	5.0	5.1	4.2	3.8	4.2	3.9	3.8	3.9	5.6	5.4
30	30	n.a.	n.a.	n.a.	n.a.	n.a.	2.7	1.7	4.2	4.2	6.1	5.0	7.1	6.1	6.6	4.2	4.1
	60	64.3	5.0	95.0	69.9	n.a.	3.9	2.6	4.6	4.2	6.5	5.3	6.6	6.5	6.4	4.5	4.6
	120	46.7	5.5	65.3	48.2	56.1	3.5	4.4	5.3	5.6	5.5	4.4	5.7	5.5	5.3	5.4	5.5
	240	27.6	5.2	37.7	28.1	28.3	3.3	4.3	5.5	4.9	4.6	5.2	4.7	4.6	4.7	5.1	3.4
Under $H_1 : a_i = 0.07,\ i = 1, ..., N$																	
10	30	7.2	5.6	7.2	7.2	7.2	7.1	2.3	4.6	4.9	8.0	9.9	8.0	8.0	9.5	4.9	7.8
	60	8.6	8.2	8.6	8.6	7.9	6.1	5.6	6.3	6.2	12.0	11.8	12.0	12.0	13.1	10.4	11.2
	120	9.0	10.0	9.0	9.0	9.1	9.9	7.0	7.6	7.5	17.5	18.6	17.5	17.5	19.3	17.2	20.2
	240	11.1	11.9	11.1	11.1	14.2	8.5	10.4	12.5	13.7	29.7	28.9	29.7	29.7	31.0	32.3	37.7
20	30	5.0	8.2	5.0	5.0	n.a.	2.9	2.4	6.3	5.6	12.9	13.1	12.9	12.9	13.4	8.6	11.8
	60	7.6	9.6	7.6	7.6	7.4	4.1	4.5	7.1	7.4	15.8	18.0	15.8	15.8	19.9	14.5	18.8
	120	11.7	11.8	11.7	11.7	12.1	8.7	7.9	9.6	12.3	24.5	31.5	24.5	24.5	32.1	31.0	37.6
	240	19.2	20.9	19.2	19.2	20.7	9.6	10.6	15.9	19.7	54.1	48.7	54.1	54.1	56.2	56.7	65.8
30	30	n.a.	n.a.	n.a.	n.a.	n.a.	2.9	1.0	5.2	4.9	16.2	18.5	16.2	16.2	21.6	9.3	13.7
	60	8.7	11.5	8.7	8.7	n.a.	4.4	3.0	7.9	7.3	24.8	26.4	24.8	24.8	26.0	21.5	29.8
	120	11.3	14.9	11.3	11.3	12.3	5.1	7.0	10.1	11.4	39.9	40.8	39.9	39.9	43.2	40.1	47.8
	240	25.9	27.7	25.9	25.9	32.1	9.6	12.3	21.2	25.6	62.5	62.1	62.5	62.5	64.3	71.8	81.4

Note: The MC tests, J_2^{MC} and J_2^{*MC}, are applied assuming the form of heteroskedasticity is known. Nominal level is 0.05 and entries are percentage rates. Results based on 1000 replications. The abbreviation n.a. stands for not applicable.

Table 2.3: **Mean-variance efficiency tests of the Sharpe-Lintner version of the CAPM.**

Time	J_1	J_2^{MMC}	J_2	J_3	J_4	SB	WB	SS	WS	J_1^*	J_2^{*MMC}	J_2^*	J_3^*	J_4^*	S^*	W^*
42-year period																
1/65–12/06	1.78	[3, 5]	17.86	17.61	18.01	1.39	1.45	4.95	6.61	4.32	[2, 5]	4.31	4.29	4.56	-0.50	0.13
	(0.06)	(0.07)	(0.05)	(0.06)	(0.05)	(0.84)	(0.79)	(0.89)	(0.76)	(0.04)	(0.05)	(0.04)	(0.04)	(0.03)	(0.61)	(0.90)
5-year subperiods and a 7-year subperiod																
1/65–12/69	2.02	[3, 21]	20.68	18.27	20.21	2.56	2.97	23.33	41.31	9.43	[1, 35]	9.04	8.66	9.53	1.46	1.98
	(0.05)	(0.07)	(0.02)	(0.05)	(0.03)	(0.10)	(0.03)	(0.01)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.14)	(0.05)
1/70–12/74	2.10	[2, 5]	21.41	18.91	19.61	2.56	2.52	40.80	39.05	0.96	[1, 6]	0.98	0.94	0.87	-2.56	-2.54
	(0.04)	(0.04)	(0.02)	(0.04)	(0.03)	(0.10)	(0.11)	(0.00)	(0.00)	(0.33)	(0.38)	(0.32)	(0.33)	(0.35)	(0.01)	(0.01)
1/75–12/79	1.94	[2, 5]	20.03	17.69	26.20	3.29	3.55	24.93	44.94	4.82	[1, 10]	4.79	4.59	5.46	1.10	1.90
	(0.06)	(0.07)	(0.03)	(0.06)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.03)	(0.05)	(0.03)	(0.03)	(0.02)	(0.27)	(0.06)
1/80–12/84	1.17	[3, 23]	12.84	11.34	11.46	1.46	1.02	4.13	2.25	1.54	[1, 35]	1.57	1.51	1.58	-0.37	-0.50
	(0.33)	(0.36)	(0.23)	(0.33)	(0.32)	(0.79)	(0.98)	(0.94)	(0.99)	(0.22)	(0.25)	(0.21)	(0.22)	(0.21)	(0.72)	(0.61)
1/85–12/89	1.72	[4, 35]	18.08	15.97	15.35	1.83	1.53	10.93	7.20	3.04	[1, 35]	3.07	2.94	2.19	-1.46	-0.77
	(0.10)	(0.13)	(0.05)	(0.10)	(0.12)	(0.50)	(0.74)	(0.36)	(0.71)	(0.09)	(0.12)	(0.08)	(0.09)	(0.14)	(0.14)	(0.44)
1/90–12/94	1.09	[2, 5]	12.08	10.67	10.85	1.46	2.05	6.67	18.86	0.09	[1, 5]	0.10	0.09	0.10	0.73	1.53
	(0.39)	(0.40)	(0.28)	(0.38)	(0.37)	(0.79)	(0.34)	(0.76)	(0.04)	(0.76)	(0.84)	(0.76)	(0.76)	(0.76)	(0.46)	(0.12)
1/95–12/99	1.14	[2, 5]	12.58	11.11	12.51	2.19	1.88	8.27	9.62	1.09	[1, 35]	1.12	1.07	0.89	0.73	0.61
	(0.35)	(0.41)	(0.25)	(0.35)	(0.25)	(0.25)	(0.46)	(0.60)	(0.47)	(0.30)	(0.34)	(0.29)	(0.30)	(0.35)	(0.46)	(0.54)
1/00–12/06	1.90	[2, 5]	19.44	17.82	17.79	2.47	2.56	41.81	37.11	9.18	[1, 10]	8.91	8.65	9.01	1.23	1.56
	(0.06)	(0.05)	(0.03)	(0.06)	(0.06)	(0.13)	(0.10)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.22)	(0.12)
10-year subperiods and a 12-year subperiod																
1/65–12/74	2.33	[3, 5]	23.22	21.87	23.07	0.77	0.96	2.33	1.68	1.33	[2, 12]	1.35	1.32	1.27	0.26	-0.05
	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.99)	(0.98)	(0.99)	(0.99)	(0.25)	(0.27)	(0.25)	(0.25)	(0.26)	(0.80)	(0.96)
1/75–12/84	2.18	[2, 4]	21.87	20.60	25.51	3.10	3.20	60.67	55.14	6.82	[1, 6]	6.75	6.60	8.01	2.58	2.33
	(0.02)	(0.02)	(0.02)	(0.02)	(0.00)	(0.02)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.02)
1/85–12/94	1.90	[2, 5]	19.25	18.13	15.72	1.81	1.52	9.13	5.73	0.59	[1, 6]	0.60	0.59	0.53	0.77	0.16
	(0.05)	(0.06)	(0.04)	(0.05)	(0.11)	(0.52)	(0.75)	(0.52)	(0.84)	(0.44)	(0.58)	(0.44)	(0.44)	(0.47)	(0.44)	(0.87)
1/95–12/06	1.39	[2, 5]	14.34	13.65	13.51	3.30	2.99	36.38	21.23	3.34	[2, 14]	3.35	3.29	3.03	2.12	1.65
	(0.19)	(0.21)	(0.16)	(0.19)	(0.20)	(0.01)	(0.03)	(0.00)	(0.02)	(0.07)	(0.08)	(0.07)	(0.07)	(0.08)	(0.03)	(0.10)

Note: Test results are based on 10 size-based equity portfolios. The CRSP value-weighted index including distributions is used to proxy the market portfolio and a one-month Treasury bill is used as the riskless asset. The numbers in parentheses are p-values. The MMC tests are based on Student-t errors with unknown degrees of freedom. The square brackets are confidence intervals with level 0.975 for the degrees-of-freedom parameter and the MMC p-values are obtained by maximizing over the confidence interval.

Table 2.4: **Estimated intercept values for individual portfolios and the aggregated one.**

Time	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_A
42-year period											
1/65–12/06	7.19	3.83	2.74	2.52	1.97	2.52	2.26	1.29	1.23	-0.42	25.19
5-year subperiods and a 7-year subperiod											
1/65-12/69	26.00	19.07	12.44	12.02	10.48	7.62	6.46	4.44	3.03	-1.87	99.71
1/70-12/74	-3.82	-8.65	-7.61	-7.56	-7.09	-4.73	-3.10	-2.67	0.06	0.84	-44.35
1/75-12/79	15.46	14.22	13.64	14.29	10.89	11.63	9.56	6.31	3.68	-2.23	97.49
1/80-12/84	6.64	4.50	6.10	5.13	3.38	3.46	3.16	1.42	0.99	-0.85	33.96
1/85-12/89	-7.62	-7.52	-5.95	-5.91	-6.21	-3.90	-1.29	-0.88	-0.73	0.77	-39.27
1/90-12/94	6.64	2.35	-0.71	-0.97	1.62	1.28	0.76	0.46	1.08	-0.18	9.09
1/95-12/99	4.50	-0.96	-2.81	-4.27	-3.45	-4.78	-6.08	-6.19	-5.66	1.66	-28.05
1/00-12/06	9.65	7.39	8.59	8.85	10.01	9.48	8.28	6.88	5.88	-1.59	73.45
10-year subperiods and a 12-year subperiod											
1/65-12/74	12.27	6.18	3.29	2.96	2.33	1.99	2.14	1.13	1.60	-0.63	33.30
1/75-12/84	12.18	10.29	10.53	10.33	7.62	7.87	6.64	4.07	2.52	-1.62	70.46
1/85-12/94	-0.20	-2.43	-3.28	-3.36	-3.85	-1.25	-0.30	-0.17	0.15	0.28	-14.43
1/95-12/06	6.75	5.17	6.01	6.19	7.00	6.63	5.79	4.81	4.11	-1.11	51.39

Note: For each time period, the line entires are the estimated intercept in a regression of the portfolio excess returns on the market excess returns. The results for the individual portfolios are labeled by P_i and those for the aggregated one by $P_A = \sum_{i=1}^{10} P_i$. The intercept values are multiplied by 1000.

Table 2.5: **Covariance matrix of market model residuals.**

Portfolio	1	2	3	4	5	6	7	8	9	10
1	4.16									
2	2.83	2.28								
3	2.20	1.74	1.55							
4	1.89	1.53	1.31	1.24						
5	1.63	1.33	1.15	1.06	1.01					
6	1.31	1.07	0.94	0.87	0.82	0.75				
7	0.98	0.82	0.73	0.69	0.65	0.58	0.53			
8	0.65	0.55	0.51	0.49	0.46	0.42	0.36	0.32		
9	0.35	0.30	0.28	0.28	0.27	0.26	0.24	0.20	0.18	
10	-0.25	-0.21	-0.18	-0.17	-0.16	-0.14	-0.12	-0.09	-0.06	0.03

Note: Entire are the covariances (multiplied by 1000) of the market model residuals on 10 size-based equity portfolios from January 1965 to December 2006. The CRSP value-weighted index including distributions is used to proxy the market portfolio and a one-month Treasury bill is used as the riskless asset.

Chapter 3

Testing Linear Factor Pricing Models in Large Cross-Sections: A Distribution-Free Approach

Abstract

We develop a finite-sample procedure to test the beta-pricing representation of linear factor pricing models that is applicable even if the number of test assets is greater than the length of the time series. Further, we make no parametric assumption about the distribution of the disturbances in the factor model. Our framework leaves open the possibility of unknown forms of time-varying non-normalities, heteroskedasticity, and even outliers in the asset returns. The power of the proposed test procedure increases as either the times series lengthens or the cross-section becomes larger. Finally, we illustrate the new procedure by testing the well-known Fama-French factor model over 5-year subsamples of monthly returns on 100 U.S. equity portfolios formed on size and book-to-market.

JEL classification: C12; C14; C22; C33; G12

Keywords: CAPM; Beta pricing; Heteroskedasticity; Non-parametric tests; Robust inference

3.1 Introduction

Many asset pricing models predict that expected returns depend linearly on "beta" coefficients relative to one or more portfolios or factors. The beta is the regression coefficient of the asset return on the factor. In the capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965), the single beta measures the systematic risk or co-movement with the returns on the market portfolio. Accordingly, assets with higher betas should offer in equilibrium higher expected returns. The Arbitrage Pricing Theory (APT) of Ross (1976), developed on the basis of arbitrage arguments, can be more general than the CAPM in that it relates expected returns with multiple beta coefficients. Merton (1973) and Breeden (1979) develop models based on investor optimization and equilibrium arguments that also lead to multiple-beta pricing.

Empirical tests of the validity of beta pricing relationships are often conducted within the context of multivariate linear factor models. When the factors are traded portfolios and a riskfree asset is available, exact factor pricing implies that the vector of asset return intercepts will be zero. These tests are interpreted as tests of the mean-variance efficiency of a benchmark portfolio in the single beta model or that some combination of the factor portfolios is mean-variance efficient in multiple beta models. In this context, standard asymptotic theory provides a poor approximation to the finite-sample distribution of the usual Wald and likelihood ratio (LR) test statistics, even with fairly large samples. Shanken (1996), Campbell, Lo, and MacKinlay (1997), and Dufour and Khalaf (2002) document severe size distortions for those tests, with over rejections growing quickly when the number of equations in the multivariate model increases. The simulation evidence in Ferson and Foerster (1994) and Gungor and Luger (2009) shows that tests based on the Generalized Method of Moments (GMM) à la MacKinlay and Richardson (1991) suffer from the same problem. As a result, empirical tests of beta-pricing representations can be severely affected and can lead to spurious rejections of their validity.

The assumptions underlying standard asymptotic arguments can be questionable when dealing with financial asset returns data. In the context of the consumption CAPM, Kocherlakota (1997) shows that model errors are so heavy-tailed that they do not satisfy the Central Limit Theorem. In such an environment, standard methods of inference can lead to spurious rejections even asymptotically and Kocherlakota instead relies on jackknifing to devise a method of testing the consumption CAPM. Similarly, Affleck-Graves and McDonald (1989) and Chou and Zhou (2006) propose the use of bootstrap techniques to provide more robust and reliable asset pricing tests.

There are very few methods that provide truly exact, finite-sample tests.[1] The most prominent one is probably the F-test of Gibbons, Ross, and Shanken (1989) (GRS). The exact distribution theory for that test rests on the assumption that the

[1]A number of Bayesian approaches have also been proposed. These include Shanken (1987), Harvey and Zhou (1990), and Kandel, McCulloch, and Stambaugh (1995).

model errors are jointly normal and independent and identically distributed through time. As we already mentioned, there is ample evidence that financial returns exhibit non-normalities; for more evidence, see Fama (1965), Blattberg and Gonedes (1974), Hsu (1982), Affleck-Graves and McDonald (1989), and Zhou (1993). Beaulieu, Dufour, and Khalaf (2007) generalize the GRS approach for testing mean-variance efficiency. Their simulation-based approach does not necessarily assume normality but it does nevertheless require that the disturbance distribution be parametrically specified, at least up to a finite number of unknown nuisance parameters. Gungor and Luger (2009) propose exact tests of the mean-variance efficiency of a single reference portfolio, whose exactness does not depend on any parametric assumptions.

In this paper we extend the idea of Gungor and Luger (2009) to obtain tests of multiple-beta pricing representations that relax two assumptions of the GRS framework: (i) the normality assumption and (ii) the restriction on the number of test assets. The proposed test procedure is based on finite-sample pivots that are valid without any assumptions about the distribution of the disturbances in the factor model. We propose an adaptive approach based on a split-sample technique to obtain a single portfolio representation judiciously formed to avoid power losses that can occur in simple portfolio groupings. For other examples of split-sample techniques, see Dufour and Taamouti (2005) and Dufour and Taamouti (2010). A very attractive feature of our approach is that it is applicable even if the number of test assets is greater than the length of the time series. This stands in sharp contrast to the GRS test or any other approach based on usual estimates of the disturbance covariance matrix. In order to avoid singularities and be computable, those approaches require the size of the cross-section be less than that of the time series. In fact, great care must be taken when applying the GRS test since its power does not increase monotonically with the number of test assets and all the power may be lost if too many are included. This problem is related to the fact that the number of covariances that need to be estimated grows rapidly with the number of included test assets. As a result, the precision with which this increasing number of parameters can be estimated deteriorates given a fixed time-series.[2]

Our proposed test procedure then exploits results from Coudin and Dufour (2009) to construct confidence sets for the model parameters by inverting exact sign-based statistics. The motivation for using this technique comes from an impossibility result due to Lehmann ans Stein (1949) that shows that the *only* tests which yield reliable inference under sufficiently general distributional assumptions, allowing non-normal, possibly heteroskedastic, independent observations are based on sign statistics. This means that all other methods, including the standard heteroskedasticity and autocorrelation-corrected (HAC) methods developed by White (1980) and Newey and West (1987) among others, which are not based on signs, cannot be proved to be

[2]The notorious noisiness of unrestricted sample covariances is a well-known problem in the portfolio management literature; see Michaud (1989), Jagannathan and Ma (2003), and Ledoit and Wolf (2003, 2004), among others.

valid and reliable for any sample size.

The paper is organized as follows. Section 2 presents the linear factor model used to describe the asset returns, the null hypothesis to be tested, and the benchmark GRS test. We provide an illustration of the effects of increasing the number of test assets on the power of the GRS test. In Section 3 we develop the new test procedure. We begin by presenting the statistical framework and then proceed to describe each step of the procedure. Section 4 contains the results of simulation experiments designed to compare the performance of the proposed test procedure with several of the standard tests. We also present evidence on the robustness of our procedure to departures from the maintained assumptions. In Section 5 we apply the procedure to test the Sharpe-Lintner version of the CAPM and the well-known Fama-French three factor model. Section 6 concludes the paper.

3.2 Factor structure

Suppose there exists a riskless asset for each period of time and define $\mathbf{r}_t$ an $N \times 1$ vector of time-t returns on N assets in excess of riskless rate of return. Suppose further that those excess returns are described by the linear K-factor model

$$\mathbf{r}_t = \mathbf{a} + \mathbf{B}\mathbf{f}_t + \boldsymbol{\varepsilon}_t, \tag{3.2.1}$$

where $\mathbf{f}_t$ is a $K \times 1$ vector of common factor portfolio excess returns, $\mathbf{B}$ is the $N \times K$ matrix of betas (or factor loadings), and $\mathbf{a}$ and $\boldsymbol{\varepsilon}_t$ are $N \times 1$ vectors of factor model intercepts and disturbances, respectively. The vector $\boldsymbol{\varepsilon}_t$ is usually assumed to have well-defined first and second moments satisfying $E[\boldsymbol{\varepsilon}_t|\mathbf{f}_t] = \mathbf{0}$ and $E[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'|\mathbf{f}_t] = \boldsymbol{\Sigma}$, a finite $N \times N$ matrix.

Exact factor pricing implies that expected returns depend linearly on the betas associated to the factor portfolio returns:

$$E_t[\mathbf{r}_t] = \mathbf{B}\boldsymbol{\lambda}_K, \tag{3.2.2}$$

where $\boldsymbol{\lambda}_K$ is a $K \times 1$ vector of expected excess returns associated with $\mathbf{f}_t$, which represent market-wide risk premiums since they apply to all traded securities. The beta-pricing representation in (3.2.2) is a generalization of the CAPM of Sharpe (1964) and Lintner (1965), which asserts that the expected excess return on an asset is linearly related to its single beta, which measures the asset's systematic risk or co-movement with the excess return on the market portfolio—the portfolio of all invested wealth. Equivalently, the CAPM says that the market portfolio is mean-variance efficient in the investment universe comprising all possible assets.[3] The pricing relationship in (3.2.2) is more general since it says that a combination (portfolio) of the factor

[3]A benchmark portfolio with excess returns r_p is said to be mean-variance efficient with respect to a given set of N test assets with excess returns $\mathbf{r}_t$ if it not possible to form another portfolio of those N assets and the benchmark portfolio with the same variance as r_p but a higher expected return.

portfolios is mean-variance efficient; see Jobson (1982), Jobson and Korkie (1982, 1985), Grinblatt and Titman (1987), Shanken (1987), and Huberman, Kandel, and Stambaugh (1987) for more on the relation between factor models and mean-variance efficiency.

The beta-pricing representation in (3.2.2) is a restriction on expected returns which can be assessed by testing the hypothesis

$$H_0 : \mathbf{a} = \mathbf{0} \tag{3.2.3}$$

under the maintained factor structure specification in (3.2.1). If the pricing errors, $\mathbf{a}$, are in fact different from zero, then (3.2.2) does not hold meaning that there is no way to combine the factor portfolios to obtain one that is mean-variance efficient.

Gibbons, Ross, and Shanken (1989) (GRS) propose a multivariate F-test of (3.2.3) that all the pricing errors are jointly equal to zero. Their test assumes that the vectors of disturbance terms $\boldsymbol{\varepsilon}_t$, $t = 1, ..., T$, in (3.2.1) are independent and normally distributed around zero with non-singular cross-sectional covariance matrix each period, conditional on the the $T \times K$ collection of factors $\mathbf{F} = [\boldsymbol{f}'_1, ..., \boldsymbol{f}'_T]'$. Under normality, the methods of maximum likelihood and ordinary least squares (OLS) yield the same unconstrained estimates of $\mathbf{a}$ and $\mathbf{B}$:

$$\hat{\mathbf{a}} = \bar{\mathbf{r}} - \hat{\mathbf{B}}\bar{\mathbf{f}}, \tag{3.2.4}$$

$$\hat{\mathbf{B}} = \left[\sum_{t=1}^{T}(\mathbf{r}_t - \bar{\mathbf{r}})(\mathbf{f}_t - \bar{\mathbf{f}})'\right]\left[\sum_{t=1}^{T}(\mathbf{f}_t - \bar{\mathbf{f}})(\mathbf{f}_t - \bar{\mathbf{f}})'\right]^{-1}, \tag{3.2.5}$$

where $\bar{\mathbf{r}} = T^{-1}\sum_{t=1}^{T}\mathbf{r}_t$ and $\bar{\mathbf{f}} = T^{-1}\sum_{t=1}^{T}\mathbf{f}_t$, and the estimate of the disturbance covariance matrix is

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{r}_t - \hat{\mathbf{a}} - \hat{\mathbf{B}}\mathbf{f}_t)(\mathbf{r}_t - \hat{\mathbf{a}} - \hat{\mathbf{B}}\mathbf{f}_t)'. \tag{3.2.6}$$

The GRS test statistic is

$$J_1 = \frac{T-N-K}{N}\left[1 + \bar{\mathbf{f}}'\hat{\boldsymbol{\Omega}}^{-1}\bar{\mathbf{f}}\right]^{-1}\hat{\mathbf{a}}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\mathbf{a}}, \tag{3.2.7}$$

where $\hat{\boldsymbol{\Omega}}$ is given by

$$\hat{\boldsymbol{\Omega}} = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{f}_t - \bar{\mathbf{f}})(\mathbf{f}_t - \bar{\mathbf{f}})'.$$

Under the null hypothesis H_0, the statistic J_1 follows a central F distribution with N degrees of freedom in the numerator and $(T-N-K)$ degrees of freedom in the denominator.

In practical applications of the GRS test, one needs to decide the appropriate number N of test assets to include. It might seem natural to try to use as many test assets as possible in order to increase the probability of rejecting H_0 when it is false. As the test asset universe expands it becomes more likely that non-zero pricing errors will be detected, if indeed there are any. However, the choice of N is restricted by T in order to keep the estimate of the disturbance covariance matrix in (3.2.6) from becoming singular, and the choice of T itself is often restricted owing to concerns about parameter stability. For instance, it is quite common to see studies where $T = 60$ monthly returns and N is between 10 and 30. The effects of increasing the number of test assets on test power is discussed in GRS and Sentana (2009). When N increases, three effects come into play: (i) the increase in the value of J_1's non-centrality parameter, which increases power, (ii) the increase in the number of degrees of freedom in the numerator, which decreases power, and (iii) the decrease in the number of degrees of freedom in the denominator, which also decreases power.

To illustrate the net effect of increasing N on the power of the GRS test, we simulated model (3.2.1) with $K = 1$, where the returns on the single factor are random draws from the standard normal distribution. The elements of the independent disturbance vector were also drawn from the standard normal distribution thereby ensuring the exactness of the GRS test. We set $T = 60$ and considered $a_i = 0.05$, 0.10, and $a_i = 0.15$ for $i = 1, ..., N$ and we let the number of test assets N range from 1 to 58. Figure 1 shows the power of the GRS test as a function of N, where for any given N the higher power is associated with higher pricing errors. In line with the discussion in GRS, this figure clearly shows the power of the test given this specification rising as N increases up to about one half of T and then decreasing beyond that. So great care must be taken when choosing the number of test assets since power does not increase monotonically with N and if the cross-section is too large, then the GRS test may lose all its power or may not even be computable. In fact, any procedure that relies on standard unrestricted estimates of the covariance matrix of regression disturbances will have this singularity problem when N exceeds T.

3.3 Test procedure

In this section we develop a procedure to test H_0 in (3.2.3) that relaxes three assumptions of the GRS test: (i) the assumption of identically distributed disturbances, (ii) the assumption of normally distributed disturbances, and (iii) the restriction on the number of test assets. Our approach is motivated by results from classical non-parametric statistics that show that the *only* tests which yield reliable inference under sufficiently general distributional assumptions, allowing non-normal, possibly heteroskedastic, independent observations are ones that are conditional on the absolute values of the observations; i.e., they must be based on sign statistics. This result is due to Lehmann and Stein (1949); see also Pratt and Gibbons (1981, p. 218), Dufour and Hallin (1991), and Dufour (2003). Next we present the statistical framework and then proceed to describe each step of the procedure.

3.3.1 Statistical framework

As in the GRS framework, we assume that the disturbance vectors $\boldsymbol{\varepsilon}_t$ in (3.2.1) are independently distributed over time, conditional on $\mathbf{F}$. We do not require the disturbance vectors to be identically distributed, but we do assume that they remain symmetrically distributed each period.In what follows the symbol $\stackrel{d}{=}$ stands for the equality in distribution.

Assumption 1. *The cross-sectional disturbance vectors $\boldsymbol{\varepsilon}_t$, $t = 1, ..., T$, are mutually independent, continuous, and diagonally symmetric so that $\boldsymbol{\varepsilon}_t \overset{d}{=} -\boldsymbol{\varepsilon}_t$, conditional on* $\mathbf{F}$.

The diagonal (or reflective) symmetry condition in Assumption 1 can be equivalently expressed in terms of the density function as $f(\boldsymbol{\varepsilon}_t) = f(-\boldsymbol{\varepsilon}_t)$. Recall that a random variable v is symmetric around zero if and only if $v \overset{d}{=} -v$, so the symmetry assumption made here represents the most direct non-parametric extension of univariate symmetry. See Serfling (2006) for more on multivariate symmetry. The class of distributions encompassed by the diagonal symmetry condition includes elliptically symmetric distributions, which play a very important role in mean-variance analysis because they guarantee full compatibility with expected utility maximization regardless of investor preferences; see Chamberlain (1983), Owen and Rabinovitch (1983), and Berk (1997). A random vector $\mathbf{V}$ is elliptically symmetric around the origin if its density function can be expressed as $|\boldsymbol{\Sigma}|^{-1/2} g(\mathbf{V}'\boldsymbol{\Sigma}^{-1}\mathbf{V})$ for some nonnegative scalar function $g(\cdot)$, where $\boldsymbol{\Sigma}$ is (proportional to) the covariance matrix. The class of elliptically symmetric distributions includes the well-known multivariate normal and Student-t distributions, among others. It is important to emphasize that the diagonal symmetry condition in Assumption 1 is less stringent than elliptical symmetry. For example, a mixture (finite or not) of distributions each one elliptically symmetric around the origin is not necessarily elliptically symmetric but it is diagonally symmetric. Note also that the distribution of $\mathbf{f}_t$ in (3.2.1) may be skewed thereby inducing asymmetry in the unconditional distribution of $\mathbf{r}_t$.

Assumption 1 does not require the vectors $\boldsymbol{\varepsilon}_t$ to be identically distributed nor does it restrict their degree of heterogeneity. This is a very attractive feature since it is well known that financial returns often depart quite dramatically from Gaussian conditions; see Fama (1965), Blattberg and Gonedes (1974), and Hsu (1982). In particular, the distribution of asset returns appears to have much heavier tails and is more peaked than a normal distribution. The present framework thus leaves open the possibility of unknown forms of non-normality and even heteroskedasticity. For example, when $(\mathbf{r}_t, \boldsymbol{f}_t)$ are elliptically distributed but nonnormal, the conditional covariance matrix of $\boldsymbol{\varepsilon}_t$ depends on the contemporaneous $\boldsymbol{f}_t$; see MacKinlay and Richardson (1991) and Zhou (1993). Here the covariance structure of the disturbance terms could be any function of the common factors (contemporaneous or not). The simulation study below includes a contemporaneous heteroskedasticity specification.

3.3.2 Portfolio formation

A usual practice in the application of the GRS test is to base it on portfolio groupings in order to have N much less than T. As Shanken (1996) notes, this has the potential effect of reducing the residual variances and increasing the precision with which $\mathbf{a} =$

$(a_1, ..., a_N)'$ is estimated. On the other hand, as Roll (1979) points out, individual stock expected return deviations under the alternative can cancel out in portfolios, which would reduce power. Ideally, all the pricing errors in $\mathbf{a}$ would be of the same sign to avoid power losses when forming portfolios. Our approach here is an adaptive one based on a split-sample technique, where the first subsample is used to obtain an estimate of $\mathbf{a}$. That estimate is then used to form a single portfolio that judiciously avoids power losses. Finally, a conditional test of H_0 is performed using only the returns on that portfolio observed over the second subsample. It is important to note that in the present framework this approach does not introduce any of the data-snooping size distortions (i.e. the appearance of statistical significance when the null hypothesis is true) discussed in Lo and MacKinlay (1990), since the estimation results are conditionally (on the factors) independent of the second subsample test outcomes.

Let $T = T_1 + T_2$. In matrix form, the first T_1 returns on asset i can be represented by

$$\mathbf{r}_i^1 = a_i \boldsymbol{\iota} + \mathbf{F}^1 \mathbf{b}_i + \boldsymbol{\varepsilon}_i^1, \tag{3.3.1}$$

where $\mathbf{r}_i^1 = [r_{i1}, ..., r_{iT_1}]'$ collects the time series of T_1 returns on asset i, $\boldsymbol{\iota}$ is a $T_1 \times 1$ vector of ones, $\mathbf{b}_i'$ is the i^{th} row of $\mathbf{B}$ in (3.2.1), and $\boldsymbol{\varepsilon}_i^1 = [\varepsilon_{i1}, ..., \varepsilon_{iT_1}]'$.

Assumption 2. *Only the first T_1 observations on $\mathbf{r}_t$ and $\mathbf{f}_t$ are used to compute the subsample estimates $\hat{a}_1, ..., \hat{a}_N$.*

This assumption does not restrict the choice of estimation method, so the subsample estimates $\hat{a}_1, ..., \hat{a}_N$ could be obtained by OLS, quasi-maximum likelihood, or any other method. A well-known problem with OLS is that it is very sensitive to the presence of large disturbances and outliers. An alternative estimation method is to minimize the sum of the absolute deviations in computing the regression lines (Koenker and Bassett 1978). The resulting least absolute deviations (LAD) estimator may be more efficient than OLS in heavy-tailed samples where extreme observations are more likely to occur. For more on the efficiency of LAD versus OLS, see Glahe and Hunt (1970), Hunt, Dowling, and Glahe (1974), Pfaffenberger and Dinkel (1978), Rosenberg and Carlson (1977), and Mitra (1987). The results reported below in the simulation study and the empirical application are based on LAD.

With the estimates $\hat{a}_1, ..., \hat{a}_N$ in hand, a vector of statistically motivated "portfolio" weights $\hat{\boldsymbol{\omega}} = (\hat{\omega}_1, \hat{\omega}_2, ..., \hat{\omega}_N)$ is computed according to:

$$\hat{\omega}_i = \frac{\hat{a}_i}{|\hat{a}_1| + ... + |\hat{a}_N|} = sign(\hat{a}_i)\frac{|\hat{a}_i|}{|\hat{a}_1| + ... + |\hat{a}_N|}, \tag{3.3.2}$$

for $i = 1, ..., N$, and these weights are then used to find the T_2 returns of a portfolio computed as $y_t = \sum_i^N \hat{\omega}_i r_{it}$, $t = T_1 + 1, ..., T$. Note that having a zero denominator in (3.3.2) is a zero probability event in finite samples when the disturbance terms are of the continuous type (as in Assumption 1). Let δ denote the sum of the weighted a_i's and set $\mathbf{x}_t = \mathbf{f}_t$.

Proposition 1. *Under H_0 and when Assumptions 1 and 2 hold, y_t is represented by the single equation*

$$y_t = \delta + \mathbf{x}_t'\boldsymbol{\beta} + u_t, \text{ for } t = T_1 + 1, ..., T, \tag{3.3.3}$$

where $\delta = 0$ and $(u_{T_1+1}, ..., u_T) \stackrel{d}{=} (\pm u_{T_1+1}, ..., \pm u_T)$, conditional on $\mathbf{F}$ and $\hat{\boldsymbol{\omega}}$.

Proof. The conditional expectation part of (3.3.3) follows from the common factor structure in (3.2.1), and the fact that δ is zero under H_0 is obvious. The independence of the disturbance vectors maintained in Assumption 1 implies that the T_2 vectors $\boldsymbol{\varepsilon}_t$, $t = T_1+1, ..., T$, are conditionally independent of the vector of weights $(\hat{\omega}_1, \hat{\omega}_2, ..., \hat{\omega}_N)$ given $\mathbf{F}$, since under Assumption 2 those weights are based only on the first T_1 observations of $\mathbf{r}_t$ and $\mathbf{f}_t$. Thus we see that, given $\mathbf{F}$ and $\hat{\boldsymbol{\omega}}$,

$$\big(\hat{\omega}_1\varepsilon_{1t}, \hat{\omega}_2\varepsilon_{2t}, ..., \hat{\omega}_N\varepsilon_{Nt}\big) \stackrel{d}{=} \big(- \hat{\omega}_1\varepsilon_{1t}, -\hat{\omega}_2\varepsilon_{2t}, ..., -\hat{\omega}_N\varepsilon_{Nt}\big), \tag{3.3.4}$$

for $t = T_1 + 1, ..., T$. Let $u_t = \sum_i^N \hat{\omega}_i\varepsilon_{it}$. For a given t, (3.3.4) implies that $u_t \stackrel{d}{=} -u_t$, since any linear combination of the elements of a diagonally symmetric vector is itself symmetric (Behboodian 1990, Theorem 2). Moreover, this fact applies to each of the T_2 conditionally independent random variables $u_{T_1+1}, ..., u_T$. So, given $\mathbf{F}$ and $\hat{\boldsymbol{\omega}}$, the 2^{T_2} possible T_2 vectors

$$\big(\pm |u_{T_1+1}|, \pm|u_{T_1+2}|, ..., \pm|u_T|\big)$$

are equally likely values for $(u_{T_1+1}, ..., u_T)$, where $\pm|u_t|$ means that $|u_t|$ is assigned either a positive or negative sign with probability 1/2. $\square$

The construction of a test based on a single portfolio grouping is reminiscent of a mean-variance efficiency test proposed in Bossaerts and Hillion (1997) based on $\sum_{i=1}^{N} \hat{a}_i$ and another one proposed in Gungor and Luger (2009) based on $\sum_{i=1}^{N} a_i$. Those approaches can suffer power losses depending on whenever the a_i's tend to cancel out. Splitting the sample and applying the weights in (3.3.2) when forming the portfolio offsets that problem. Note that these weights do not correspond to any of the usual ones in mean-variance analysis since finding those requires an estimate of the covariance structure and that is precisely what we are trying to avoid. Furthermore, such estimates are not meaningful in our non-parametric context where possible forms of heterogeneity are left completely unspecified. To see why the weights in (3.3.2) are reasonable, note that the sign component in the definition of $\hat{\omega}_i$ makes it more likely that all the intercept values in the equation describing $\hat{\omega}_i r_{it}$ will be positive under the alternative hypothesis. The component in (3.3.2) pertaining to the absolute values serves to give relatively more weight to the assets that seem to depart more from H_0 and to down weight those that seem to offer relatively less evidence against the null hypothesis.

3.3.3 Confidence sets

The model in (3.3.3) can be represented in matrix form as $\mathbf{y} = \delta + \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, where the elements of $\mathbf{u}$ follow what Coudin and Dufour (2009) call a strict conditional mediangale. Define a sign function as $s[x] = 1$ if $x > 0$, and $s[x] = -1$ if $x \leq 0$. The following result is an immediate consequence of the mediangale property.

Proposition 2. *Under Assumptions 1 and 2, the T_2 disturbance sign vector*

$$s(\mathbf{y} - \delta - \mathbf{X}\boldsymbol{\beta}) = \big(s[y_{T_1+1} - \delta - \mathbf{x}'_{T_1+1}\boldsymbol{\beta}], ..., s[y_T - \delta - \mathbf{x}'_T\boldsymbol{\beta}]\big)$$

follows a distribution free of nuisance parameters, conditional on $\mathbf{F}$ *and* $\hat{\boldsymbol{\omega}}$. *Its distribution can be simulated to any degree of accuracy simply by repeatedly drawing* $\tilde{S}_{T_2} = (\tilde{s}_1, ..., \tilde{s}_{T_2})$, *whose elements are independent Bernoulli variables such that* $\Pr[\tilde{s}_t = 1] = \Pr[\tilde{s}_t = -1] = 1/2$.

A corollary of this proposition is that any function of the form $\Psi = \Psi(s(\mathbf{y} - \delta - \mathbf{X}\boldsymbol{\beta}); \mathbf{F})$ is also free of nuisance parameters (i.e. pivotal), conditional on $\mathbf{F}$. To see the usefulness of this result, consider the problem of testing $H_0(\delta_0, \boldsymbol{\beta}_0) : \delta = \delta_0, \boldsymbol{\beta} = \boldsymbol{\beta}_0$ against $H_1(\delta_0, \boldsymbol{\beta}_0) : \delta \neq \delta_0$ or $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$. Under $H_0(\delta_0, \boldsymbol{\beta}_0)$, the statistic $\Psi(s(\mathbf{y} - \delta - \mathbf{X}\boldsymbol{\beta}); \mathbf{F})$ is distributed like $\Psi(\tilde{S}_{T_2}; \mathbf{F})$, conditional on $\mathbf{F}$. Suppose that $\Psi(\cdot)$ is a non-negative function. The decision rule is then to reject $H_0(\delta_0, \boldsymbol{\beta}_0)$ at level α if $\Psi(s(\mathbf{y} - \delta - \mathbf{X}\boldsymbol{\beta}); \mathbf{F})$ is greater than the $(1 - \alpha)$-quantile of the simulated distribution of $\Psi(\tilde{S}_{T_2}; \mathbf{F})$.

Following Coudin and Dufour (2009), we consider two test statistics given by the quadratic forms

$$SX(\delta_0, \boldsymbol{\beta}_0) = s(\mathbf{y} - \delta_0 - \mathbf{X}\boldsymbol{\beta}_0)'\mathbf{X}\mathbf{X}'s(\mathbf{y} - \delta_0 - \mathbf{X}\boldsymbol{\beta}_0), \tag{3.3.5}$$

$$SP(\delta_0, \boldsymbol{\beta}_0) = s(\mathbf{y} - \delta_0 - \mathbf{X}\boldsymbol{\beta}_0)'\mathbf{P}(\mathbf{X})s(\mathbf{y} - \delta_0 - \mathbf{X}\boldsymbol{\beta}_0), \tag{3.3.6}$$

where $\mathbf{P}(\mathbf{X}) = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ projects orthogonally onto the subspace spanned by the columns of $\mathbf{X}$. Boldin, Simonova, and Tyurin (1997) show that these statistics can be associated with locally most powerful tests in the case of i.i.d. disturbances under some regularity conditions. Coudin and Dufour extend that proof to disturbances that satisfy the mediangale property. It is interesting to note that (3.3.6) can be interpreted as a sign analogue of the F test for testing the hypothesis that all the coefficients in a regression of $s(\mathbf{y} - \delta_0 - \mathbf{X}\boldsymbol{\beta}_0)$ on $\mathbf{X}$ are zero.

An exactly distribution-free confidence set for δ and $\boldsymbol{\beta}$ can be constructed simply by inverting either (3.3.5) or (3.3.6). Consider the test statistic in (3.3.6) for example, and let c_α represent its one-sided α-level simulated critical value. A joint confidence set, say $C_{1-\alpha}(\delta, \boldsymbol{\beta})$, with level $1-\alpha$ for δ and $\boldsymbol{\beta}$ is simply the collection of all values of $\delta_0, \boldsymbol{\beta}_0$ for which $SP(\delta_0, \boldsymbol{\beta}_0)$ is less than c_α. Note that the critical value c_α only needs to be computed once, since it does not depend on $\delta_0, \boldsymbol{\beta}_0$.

From the joint confidence set, it is possible to derive conservative confidence sets and intervals for general functions of the parameters $\delta, \boldsymbol{\beta}$ using the projection method in Coudin and Dufour (2009); see also Abdelkhalek and Dufour (1998), Dufour and Jasiak (2001), and Dufour and Taamouti (2005) for other examples of this technique. Consider a non-linear function $g(\delta, \boldsymbol{\beta})$ of $\delta, \boldsymbol{\beta}$. It is easy to see that $(\delta, \boldsymbol{\beta}) \in C_{1-\alpha}(\delta, \boldsymbol{\beta}) \Rightarrow g(\delta, \boldsymbol{\beta}) \in g(C_{1-\alpha}(\delta, \boldsymbol{\beta}))$ so that $\Pr[(\delta, \boldsymbol{\beta}) \in C_{1-\alpha}(\delta, \boldsymbol{\beta})] \geq 1 - \alpha \Rightarrow \Pr[g(\delta, \boldsymbol{\beta}) \in g(C_{1-\alpha}(\delta, \boldsymbol{\beta}))] \geq 1 - \alpha$. This means that $g(C_{1-\alpha}(\delta, \boldsymbol{\beta}))$ is a conservative confidence set for $g(\delta, \boldsymbol{\beta})$; i.e., one for which the level is at least $1 - \alpha$. In the special case when $g(\delta, \boldsymbol{\beta})$ is scalar, the interval

$$\left[\inf_{(\delta_0, \boldsymbol{\beta}_0) \in C_{1-\alpha}(\delta, \boldsymbol{\beta})} g(\delta_0, \boldsymbol{\beta}_0), \sup_{(\delta_0, \boldsymbol{\beta}_0) \in C_{1-\alpha}(\delta, \boldsymbol{\beta})} g(\delta_0, \boldsymbol{\beta}_0)\right]$$

satisfies

$$\Pr\left[\inf_{(\delta_0, \boldsymbol{\beta}_0) \in C_{1-\alpha}(\delta, \boldsymbol{\beta})} g(\delta_0, \boldsymbol{\beta}_0) \leq g(\delta, \boldsymbol{\beta}) \leq \sup_{(\delta_0, \boldsymbol{\beta}_0) \in C_{1-\alpha}(\delta, \boldsymbol{\beta})} g(\delta_0, \boldsymbol{\beta}_0)\right] \geq 1 - \alpha.$$

Hence, a confidence interval of the form $[\hat{\delta}_L, \hat{\delta}_U]$ for δ in model (3.3.3) can be found as

$$\begin{aligned} \hat{\delta}_L = & \operatorname*{argmin}_{(\delta_0, \boldsymbol{\beta}_0) \in \mathbb{R} \times \mathbb{R}^K} \delta_0, & \hat{\delta}_U = & \operatorname*{argmax}_{(\delta_0, \boldsymbol{\beta}_0) \in \mathbb{R} \times \mathbb{R}^K} \delta_0, \\ & \text{subject to} \quad SP(\delta_0, \boldsymbol{\beta}_0) < c_\alpha, & & \text{subject to} \quad SP(\delta_0, \boldsymbol{\beta}_0) < c_\alpha. \end{aligned} \tag{3.3.7}$$

Once the solutions in (3.3.7) are found, the null hypothesis $H_0 : \mathbf{a} = \mathbf{0}$ in (3.2.3) is rejected at level α if zero is not contained in $[\hat{\delta}_L, \hat{\delta}_U]$, otherwise there is not sufficient evidence to reject it at that level of significance.

Searching over the $\mathbb{R} \times \mathbb{R}^K$ domain in (3.3.7) is obviously not practical and some restrictions need to be imposed. Here we perform that step by specifying a fine grid of relevant points $\mathcal{B}(\hat{\delta}_0, \hat{\boldsymbol{\beta}}_0)$ around $\hat{\delta}_0, \hat{\boldsymbol{\beta}}_0$ and calculate $SP(\delta_0, \boldsymbol{\beta}_0)$ at each of those points.[4] An important remark about computation is that a single pass over the grid is enough to establish both the joint confidence set and the limits of the marginal confidence interval for δ. Note also that the grid search can be stopped and the null hypothesis can no longer be rejected at the α level as soon as zero gets included in the marginal confidence interval for δ.

3.3.4 Summary of test procedure

Suppose that one wishes to use the SP statistic in (3.3.6). In a preliminary step, the distribution of that statistic is simulated to the desired degree of accuracy and a one-sided α-level critical value, c_α, is determined. The rest of the test procedure then proceeds according to the following steps.

1. The estimates $\hat{a}_i$ of a_i, $i = 1, ..., N$, are computed using the first subsample of observations, r_{it} and $\mathbf{f}_t$, $t = 1, ..., T_1$.

[4] More sophisticated global optimization methods could be used to solve for the limits of the marginal confidence interval. For instance, Coudin and Dufour (2009) make use of a simulated annealing algorithm (Goffe, Ferrier, and Rogers 1994). The advantage of the grid search is that it is completely reliable.

2. For each $i = 1, ..., N$, weights $\hat{\omega}_i$ are computed according to:

$$\hat{\omega}_i = \frac{\hat{a}_i}{|\hat{a}_1| + ... + |\hat{a}_N|},$$

and T_2 returns of a portfolio are computed as $y_t = \sum_i^N \hat{\omega}_i r_{it}$, $t = T_1 + 1, ..., T$.

3. For each candidate point $(\delta_0, \boldsymbol{\beta}_0) \in \mathcal{B}(\hat{\delta}_0, \hat{\boldsymbol{\beta}}_0)$, the statistic $SP(\delta_0, \boldsymbol{\beta}_0)$ is computed. The limits of the marginal confidence interval, $\hat{\delta}_L$ and $\hat{\delta}_U$, are found as:

$$\hat{\delta}_L = \underset{(\delta_0,\boldsymbol{\beta}_0)\in\mathcal{B}(\hat{\delta}_0,\hat{\boldsymbol{\beta}}_0)}{\text{argmin}} \quad \delta_0, \qquad \hat{\delta}_U = \underset{(\delta_0,\boldsymbol{\beta}_0)\in\mathcal{B}(\hat{\delta}_0,\hat{\boldsymbol{\beta}}_0)}{\text{argmax}} \quad \delta_0,$$

$$\text{subject to} \quad SP(\delta_0, \boldsymbol{\beta}_0) < c_\alpha, \qquad \text{subject to} \quad SP(\delta_0, \boldsymbol{\beta}_0) < c_\alpha.$$

4. The null hypothesis $H_0 : \mathbf{a} = \mathbf{0}$ is rejected if $0 \notin [\hat{\delta}_L, \hat{\delta}_U]$, otherwise it is accepted.

This procedure yields an exactly distribution-free test of H_0 over the class of all disturbance distributions satisfying Assumption 1.

3.4 Simulation results

We present the results of some small-scale simulation experiments to compare the performance of the proposed test procedure with several standard tests. We also present some evidence of the robustness of the new procedure to the presence of cross-sectional error correlation. The first of the benchmarks for comparison purposes is the GRS test in (3.2.7). The other benchmarks are the usual likelihood ratio (LR) test, an adjusted LR test, and a test based on the Generalized Method of Moments (GMM). The latter is a particularly important benchmark here, since in principle it is "robust" to non-normality and heteroskedasticity of returns.

The LR test is based on a comparison of the constrained and unconstrained log-likelihood functions evaluated at the maximum likelihood estimates. The unconstrained estimates are given in (3.2.4), (3.2.5), and (3.2.6). For the constrained case, the maximum likelihood estimates are

$$
\begin{aligned}
\hat{\mathbf{B}}^* &= \left[\sum_{t=1}^{T} \mathbf{r}_t \mathbf{f}_t'\right] \left[\sum_{t=1}^{T} \mathbf{f}_t \mathbf{f}_t'\right]^{-1}, \\
\hat{\boldsymbol{\Sigma}}^* &= \frac{1}{T} \sum_{t=1}^{T} \left(\mathbf{r}_t - \hat{\mathbf{B}}^* \mathbf{f}_t\right) \left(\mathbf{r}_t - \hat{\mathbf{B}}^* \mathbf{f}_t\right)'.
\end{aligned}
$$

The LR test statistic, J_2, is then given by

$$
J_2 = T \left[\log |\hat{\boldsymbol{\Sigma}}^*| - \log |\hat{\boldsymbol{\Sigma}}|\right],
$$

which, under the null hypothesis, follows an asymptotic chi-square distribution with N degrees of freedom, χ^2_N. As we shall see, the finite sample behavior of J_2 can differ vastly from what asymptotic theory predicts. Jobson and Korkie (1982) suggest an adjustment to J_2 in order to improve its finite-sample size properties when used with critical values from the χ^2_N distribution. The adjusted statistic is

$$
J_3 = \frac{T - (N/2) - K - 1}{T} J_2,
$$

which also follows the asymptotic χ^2_N distribution, under H_0.

MacKinlay and Richardson (1991) develop tests of mean-variance efficiency in a GMM framework. For the asset pricing model in (3.2.1), the GMM tests are based on the moments of the following $(K+1)N \times 1$ vector:

$$\mathbf{g}_t(\boldsymbol{\theta}) = \begin{pmatrix} 1 \\ \\ \mathbf{f}_t \end{pmatrix} \otimes \boldsymbol{\varepsilon}_t(\boldsymbol{\theta}), \tag{3.4.1}$$

where $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) = \mathbf{r}_t - \mathbf{a} - \mathbf{B}\mathbf{f}_t$. The symbol $\otimes$ refers to the Kronecker product. Here $\boldsymbol{\theta} = (\mathbf{a}', vec(\mathbf{B})')'$, where $vec(\mathbf{B})$ is an $NK \times 1$ vector obtained by stacking the columns of $\mathbf{B}$, one below the other, with the columns ordered from left to right. The model specification in (3.2.1) implies the moment conditions $E(\mathbf{g}_t(\boldsymbol{\theta}_0)) = 0$, where $\boldsymbol{\theta}_0$ is the true parameter vector. The system in (3.4.1) is exactly identified which implies that the GMM procedure yields the same estimates of $\boldsymbol{\theta}$ as does OLS applied equation by equation. The covariance matrix of the GMM estimator $\hat{\boldsymbol{\theta}}$ is given by $\mathbf{V} = [\mathbf{D}_0'\mathbf{S}_0^{-1}\mathbf{D}_0]^{-1}$, where $\mathbf{D}_0 = E[\partial \mathbf{g}_t(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}_0']$ and $\mathbf{S}_0 = \sum_{s=-\infty}^{+\infty} E[\mathbf{g}_t(\boldsymbol{\theta}_0)\mathbf{g}_{t-s}(\boldsymbol{\theta}_0)']$; see Campbell, Lo, and MacKinlay (1997, Chapter 5). The GMM-based Wald test statistic is

$$J_4 = T\hat{\mathbf{a}}' \left[\mathbf{R}\left(\hat{\mathbf{D}}'\hat{\mathbf{S}}^{-1}\hat{\mathbf{D}}\right)^{-1}\mathbf{R}'\right]^{-1}\hat{\mathbf{a}}, \tag{3.4.2}$$

where $\hat{\mathbf{D}}$ and $\hat{\mathbf{S}}$ are consistent estimators of $\mathbf{D}_0$ and $\mathbf{S}_0$, respectively, and $\mathbf{R} = (1, \mathbf{0}_K) \otimes \mathbf{I}_N$, with $\mathbf{0}_K$ denoting a row vector of K zeros and $\mathbf{I}_N$ as the $N \times N$ identity matrix. Note that the J_4 statistic cannot be computed whenever $(K+1)N$ exceeds T, since $\hat{\mathbf{S}}$ then becomes singular.

Our implementation of the proposed test procedure is computationally intensive owing to the numerical grid search we perform in Step 3. This is not overly costly for a single application of the procedure, but it does become prohibitive for a simulation study. For that reason, we restrict our attention to cases with $K = 1$ in model (3.2.1). For convenience, the single-factor specification is given again here as

$$r_{it} = a_i + b_i f_t + \varepsilon_{it}, \quad t = 1, ..., T,\ i = 1, ..., N, \tag{3.4.3}$$

in which case the null hypothesis is a test of the mean-variance efficiency of the given portfolio. The returns of the reference portfolio, f_t, follow a stochastic volatility process:

$$f_t = \exp(h_t/2)\epsilon_t \text{ with } h_t = \lambda h_{t-1} + \xi_t,$$

where the independent terms ϵ_t and ξ_t are both i.i.d. according to a standard normal distribution and the persistence parameter λ is set to 0.5. The b_i's are randomly drawn from a uniform distribution between 0.5 and 1.5. All the tests are conducted at the nominal 5% level and critical values for $SX(\delta_0, \boldsymbol{\beta}_0))$ and $SP(\delta_0, \boldsymbol{\beta}_0)$ are determined using 10,000 simulations. In the experiments we choose mispricing values a and set half the intercept values as $a_i = a$ and the other half as $a_i = -a$. We denote this in the tables as $|a_i| = a$. The estimates of a_i, $i = 1, ..., N$, in Step 1 are found via LAD. Finally, there are 1000 replications in each experiment.

In the application of the test procedure, a choice needs to be made about where to split the sample. While this choice has no effect on the level of the tests, it obviously matters for their power. We do not have analytical results on how to split the sample, so we resort to simulations. Table 1 shows the power of the test procedure applied with the SX and SP statistics for various values of T_1/T, where $|a_i| = 0.20, 0.15$, and 0.10. Here $T = 60$ and $N = 100$ and the disturbance terms ε_{it} are drawn randomly from the Student-t distribution with ν degrees of freedom. We consider $\nu = 12$ and 6 to examine the effects of kurtosis on the power of the tests. As expected the results show that for any given value of T_1/T, the power increases as $|a_i|$ increases and decreases as the kurtosis of the disturbance terms increases. Overall, the results suggest that no less that 30% and no more than 50% of the time series observations should be used as the first subsample in order to maximize power. Accordingly, the testing strategy represented by $T_1 = 0.4T$ is pursued in the remaining comparative experiments.

We also include in our comparisons two distribution-free tests proposed by Gungor and Luger (2009) that are applicable in the single-factor context. The building block of those tests is

$$z_{it} = \left(\frac{r_{i,t+m}}{f_{t+m}} - \frac{r_{it}}{f_t} \right) \times \frac{(f_t - f_{t+m})}{f_t f_{t+m}}, \tag{3.4.4}$$

defined for $t = 1, ..., m$, where $m = T/2$ is assumed to be an integer. The first test is based on the sign statistic

$$S_i = \frac{\sum_{t=1}^{m} 0.5(s[z_{it}] + 1) - m/2}{\sqrt{m/4}} \tag{3.4.5}$$

and the second one is based the Wilcoxon signed rank statistic

$$W_i = \frac{\sum_{t=1}^{m} 0.5(s[z_{it}] + 1)\text{Rank}(|z_{it}|) - m(m+1)/4}{\sqrt{m(m+1)(2m+1)/24}}, \tag{3.4.6}$$

where $\text{Rank}(|z_{it}|)$ is the rank of $|z_{it}|$ when $|z_{i1}|, ..., |z_{im}|$ are placed in ascending order of magnitude. Gungor and Luger (2009) show that a time-series symmetry condition ensures that both (3.4.5) and (3.4.6) have limiting (as $m \to \infty$) standard normal distributions. Under the further assumption that the disturbance terms are cross-sectionally independent, conditional on $(f_1, ..., f_T)'$, their sum-type statistics

$$SD = \sum_{i=1}^{N} S_i^2 \text{ and } WD = \sum_{i=1}^{N} W_i^2 \tag{3.4.7}$$

follow an asymptotic chi-square distribution with N degrees of freedom. Simulation results show that this approximation works extremely well and just like the test procedure proposed here, the SD and WD test statistics can be calculated even if N is large.

Tables 2 and 3 show the empirical size (Panel A) and power (Panel B) when $|a_i| = 0.15$ of the considered tests for $T = 60, 120$ and $N = 10, 25, 50, 100, 125$. The power results for the J_1, J_2, J_3, and J_4 are based on size-corrected critical values, since none of those tests are exact under the two specifications we examine. It is important to emphasize that size-corrected tests are not feasible in practice, especially under the very general symmetry condition in Assumption 1. They are merely used here as theoretical benchmarks for the truly distribution-free tests. In particular, we wish to see how the power of the new tests compares to these benchmarks as T and N vary.

The results in Table 2 correspond to the single-factor model where the disturbance terms ε_{it} are i.i.d. in both the time-series and the cross-section according to a Student-t distribution with 6 degrees of freedom. From Panel A, we see that the parametric J_1 and the distribution-free SD and WD tests behave well under the null with empirical rejection rates close to the nominal level. This finding for the GRS test is in line with Affleck-Graves and McDonald (1989) who present simulation evidence showing the GRS test to be fairly robust to deviations from normality. From Table 2, the (conservative) SX and SP tests are also seen to satisfy the level constraint in the sense that the probability of a Type I error remains bounded by the nominal level of significance.[5] The J_2, J_3, and J_4 tests, however, suffer massive size distortions as the number of equations increases.[6] When $T = 120$ and $N = 100$, the LR test (J_2) rejects the true null with an empirical probability of 100% and in the case of the adjusted LR test (J_3) that probability is still above 50%. Notice as well that the J_1, J_2, and J_3 are not computable when N exceeds T, and the GMM-based J_4 cannot even be computed here as soon as $2N$ exceeds T. (Those cases are indicated with "-" in the tables.)

In Panel B of Table 2, we see the same phenomenon as in Figure 1: for a fixed T,

[5]Following the terminology in Lehmann and Romano (2005, Chapter 3), we say that a test of H_0 has *level* α if the probability of incorrectly rejecting H_0 when it is true is *not greater than* α.

[6]This overrejection problem with standard asymptotic tests in multivariate regression models is also documented in Stambaugh (1982), Jobson and Korkie (1982), Amsler and Schmidt (1985), MacKinlay (1987), Stewart (1997), and Dufour and Khalaf (2002).

the power of the GRS J_1 test rises and then eventually drops as N increases. Note that J_1, J_2, and J_3 have identical size-corrected powers, since they are all related via monotonic transformations (Campbell, Lo, and MacKinlay 1997, Chapter 5). On the contrary, the power of the SD and WD tests and that of the new SX and SP tests increases monotonically with N.

The second specification we consider resembles a stochastic volatility model and introduces dependence between the conditional covariance matrix and f_t. Specifically, we let $\varepsilon_{it} = \exp(\lambda_i f_t/2)\eta_{it}$, where the innovations η_{it} are standard normal and the λ_i's are randomly drawn from a uniform distribution between 1.5 and 2.5. It should be noted that such a contemporaneous heteroskedastic specification finds empirical support in Duffee (1995, 2001) and it is easy to see that it generates ε_{it}'s with excess kurtosis—a well-known feature of asset returns. Panel A of Table 3 reveals that all the parametric tests have massive size distortions in this case, and these over-rejections worsen as N increases for a given T.[7] When $T = 120$, the J tests all have empirical sizes around 20%. The probability of a Type I error for all those tests exceeds 65% when N is increased to 50. In sharp contrast, the four distribution-free tests satisfy the nominal 5% level constraint, no matter T and N. As in the first example, Panel B shows the power of the distribution-free tests increasing with both T and N in this heteroskedastic case.

[7] The sensitivity of the GRS test to contemporaneous heteroskedasticity is also reported in MacKinlay and Richardson (1991), Zhou (1993), and Gungor and Luger (2009).

At this point, one may wonder what is the advantage of the new SX and SP tests since the SD and WD tests of Gungor and Luger (2009) seem to have better power in Panel B of Tables 2 and 3. Those tests achieve higher power because they eliminate the b_i's from the inference problem through the long differences in (3.4.4), whereas the new tests proceed by finding set estimates of those nuisance parameters. A limitation of the SD and WD tests is that they are valid only under the assumption that the model disturbances are cross-sectionally independent. Table 4 reports the empirical size of the distribution-free tests when the cross-sectional disturbances are multivariate normal with an equicorrelation structure. Specifically, the disturbances have zero mean, unit variance, and the correlation between any two disturbances is equal to ρ, which we vary between 0.1 and 0.5. We see from Table 4 that the SD and WD tests are fairly robust to mild cross-sectional correlation, but start over-rejecting as the equicorrelation increases and this problem is further exacerbated as either T or N increases. As expected, the proposed SX and SP tests are seen to behave well in Table 4 since Assumption 1 on which they rest allows for cross-sectional disturbance correlation. The second limitation of the SD and WD tests is that they are designed for the single-factor model and cannot be easily extended to allow for multiple factors. The new SX and SP tests are illustrated next in the context of a single- and a three-factor model.

3.5 Empirical application

We illustrate the use of the new test procedure with two empirical applications. First, we present the results of the proposed tests for a single-factor asset pricing model and compare with those of the standard parametric tests and the non-parametric tests of Gungor and Luger (2009). This application investigates the mean-variance efficiency of a value-weighted market index of all stocks listed on the NYSE, AMEX, and NASDAQ. Our second application compares the results from the Fama and French (1993) three-factor model and examines whether a combination of the factor portfolios is mean-variance efficient. We use monthly value-weighted returns for the period from January 1965 to December 2009 with a total of 540 observations. Our data consist of the returns on the three Fama-French factors, one-month U.S. Treasury bill rate as the risk-free rate, and the returns on three sets of portfolios from the data library of Kenneth French. The three portfolio sets include 10 portfolios formed on size, and 25 and 100 portfolios formed on both size and book-to-market. We acknowledge that short subperiods could reduce the effects of time-variation in linear factor models. Therefore, in addition to the entire whole sample period with $T = 540$, we report the results for nine 5-year, $T = 60$, and four 10-year, $T = 120$, subperiods.

3.5.1 10 portfolios formed on size

Using the return data on 10 portfolios formed on size, we first apply the parametric J tests and the non-parametric tests to the single-factor model. The empirical results are summarized in Table 1. Columns 2–5 present the conventional J tests; columns 6 and 7 present the non-parametric SD and WD proposed by Gungor and Luger (2009); and the square brackets in columns 8 and 9 report the 95% confidence intervals for the new test procedure applied with SX and SP statistics.

Over the entire sample period, the parametric J tests tend to reject the null hypothesis with p-values no more than 1%, implying that the pricing errors from the single-factor model are jointly significantly different from zero. In contrast, the p-values for the non-parametric SD and WD, 82% and 60% respectively, and the 95% confidence intervals for the new procedure with SX and SP statistics favor non-rejection supporting the mean-variance efficiency of the market index.

In three of the nine 5-year subperiods, 1/65–12/69, 1/90–12/94, and 1/00–12/04, all the Js reject the efficiency of the market portfolio with p-values less than 8%. The results of the non-parametric SD and WD statistics are consistent with those of the J tests for 1/65–12/69 and 1/00–12/04, altough, they also reject the null in 1/70–12/74 and 1/75–12/79. The newly proposed SX and SP, however, continue to not reject the mean-variance efficiency, except for SP in the 1/75–12/79 subperiod.

In the 10-year subperiods, the results are more in line with the findings for the entire sample period, such that, the J tests tend to reject the null more often than the non-parametric tests. Out of the four subsamples, the Js indicate a rejection for the last three, while SD and WD tend to reject only in the second subperiod, and the SX and SP only in the third subperiod. Solely for the first subsample the parametric and the non-parametric tests agree on non-rejection.

The empirical results in Table 1 reveal that the inference is sensitive to the choice of the sample period as well as the choice of the test statistics. Besides the obvious difference in the parametric and non-parametric inference in the full sample, the non-parametric SD and WD, and the proposed SX and SP statistics differ markedly in the subperiods. One possible explanation for the disagreement across non-parametric tests is the existence of cross-correlations among the test portfolios. Note that, SD and WD rest on the assumption of independence while SX and SP allow for correlations across 10 portfolios, providing a more robust testing procedure for the mean-variance efficiency.

Table 2 displays the results for the Fama-French three-factor model. Columns 2–5 show the results of the parametric J tests. The square brackets in columns 6–7 show the 95% confidence intervals for the SX and SP statistics. The non-parametric SD and WD cannot be computed for large number of factors due to the reduction in the sample size caused by the long-differences required with each additional factor.

Hence, they are not included in Table red2.

For the entire 45-year sample period, the reported results for the three-factor model in Table 2 are very similar to those of the single-factor model in Table 1. The standard J tests reject the null with low p-values equal to 1% whereas the distribution-free SX and SP tests show evidence of non-rejection implying that a combination of the factor portfolios is mean-variance efficient.

In the 5-year subperiods, there is much disagreement among the parametric tests themselves. In seven of the nine subsamples, the GMM-based Wald test, J_4, tends to reject the null with p-values lower than 6%. The usual LR test, J_2, rejects the mean-variance efficiency for 1/80–12/84, 1/90–12/94, and 1/95–12/99, whereas the adjusted LR test, J_3, rejects only for 1/80–12/84 with a p-value of 6%. Note that, only in the subperiod 1/80–12/84, the GRS test, J_1, rejects the null with a p-value of 7% and agrees with the non-parametric SX and SP on non-rejection for the rest of the 5-year intervals. The results in the 10-year subperiods resemble those for the entire sample period and the J tests are consistent among each other. Over the last two subperiods, all the Js indicate a very clear rejection with p-values less than 3%, while SX and SP favor non-rejection. For the first two subperiods, however, the Fama-French asset pricing model is supported by all the test procedures.

In Table 2, the SX and SP statistics do not reject the Fama-French three-factor model in the whole period or in any of the subperiods. These non-rejections imply that the excess returns of 10 portfolios are well explained by the three factors. However, a closer look at the results in Table 1 and Table 2 clearly shows that the size and the book-to-market factors are not always required, and the market portfolio alone can price the 10 portfolios in eight of the nine 5-year and in three of the four 10-year subperiods.

Observing that the Fama-French model is never rejected by the non-parametric SX and SP statistics, one may be concerned about the ability of the new procedure to correctly reject the null, when in fact the alternative is true. The fact that the single-factor model is rejected in the subperiods 1/85-12/94 and 1/75-12/79 provides an evidence that the non-rejections by the non-parametric procedure are not due to low power.

3.5.2 25 portfolios formed on size and book-to-market

In this section, we use the return data on 25 portfolios formed on size and book-to-market. The results for the single-factor model, reported in Table 3, indicate that both the parametric J tests, and the non-parametric SD and WD strongly reject the null hypothesis of mean-variance efficiency with p-values less than 1%. The SX and SP statistics, however, are in disagreement with all the other tests by tending to not reject the null.

For 25 portfolios, the results in the 5-year and the 10-year subperiods are consistent with those of the whole sample such that, in the majority of the subperiods the null hypothesis is again rejected both by the standard J tests and the SD and WD. For the same subperiods, the confidence intervals from the SX and SP statistics indicate the exact opposite where the intercepts from the single-factor model are jointly not different from zero at the 5% significance level. In other words, SX and SP demonstrate that the excess return on the market portfolio is sufficient to explain the risk premia of the 25 portfolios.

Table 4 displays the results for the Fama-French model using 25 portfolios. In the full sample, the J tests continue to strongly reject the null with p-values less then 1%. It is, however, quite surprising to find that the three-factor model is also rejected by the SP statistic at 5% level of confidence, whereas the earlier single-factor model is not.

One drawback of the parametric J_1, J_2, J_3, and J_4 statistics is that they cannot be computed with $N \geq T$. The GMM Wald test J_4 cannot even be computed when $(K+1)N \geq T$. The matrix $\hat{S}$ is singular if this condition is not satisfied. The dashed lines in the tables denote these cases where the parametric tests cannot be calculated. None of the non-parametric tests, however, suffer from these limitations of N relative to T. Among the test statistics that are calculated, the usual LR test, J_2, tends to reject the null hypothesis in all the 5-year subperiods, except in 1/75-12/79. On the other hand, J_1 and J_3 indicate a rejection only in the subperiods 1/85-12/89 and 1/90-12/94. The 95% confidence intervals associated with SX and SP favor non-rejection in all the 5-year subperiods. The results in the 10-year subperiods are very similar to those in the 5-year, where the null is mostly rejected by the parametric tests and not rejected by the non-parametric tests.

3.5.3 100 portfolios formed on size and book-to-market

Table 5 displays the results for the single-factor model using return data on 100 portfolios formed on size and book-to-market. Similar to the earlier findings over the full sample with 25 portfolios, the J statistics decidedly reject the null with p-values close to zero. The inference of the SD and WD are in agreement with those of the parametric ones, with compatible small p-values. In sharp contrast, the 95% confidence intervals for SX and SP continue to not reject, validating the single-factor model and the mean-variance efficiency of the market portfolio.

For the 5-year subperiods, because $N = 100 > T = 60$, none of the parametric tests can be calculated. Therefore, Table 5 reports only the non-parametric results since they are free from this limitation. In six of the nine subperiods, the SD and

WD statistics tend towards a rejection where SP rejects the null only in the 1/75-12/74 subperiod. As mentioned earlier, such a difference among the non-parametric procedures is an indication of the presence of cross-correlations across 100 portfolios. In the 10-year subperiods, the J_1, J_2, and J_3 become available mostly rejecting the null with very low p-values. In the same subperiods, SX and SP indicate no rejection.

In Table 5 we present the results for the Fama-French three-factor model. In the whole 45-year period, the non-parametric SP statistic agrees with the parametric inference on rejection. Within the 5-year subsamples, the three-factor model is also rejected by the both non-parametric statistics for the 1/80-12/84 subperiod. The 10-year results resemble the previous findings, where the parametric tests in general tend to reject the null and the non-parametric ones tend not to.

3.5.4 Extreme observations

In Tables 1 through 6, we find striking difference between the parametric and non-parametric inference. In general, the parametric J tests reject both the single-factor and the Fama-French three-factor model, while the non-parametric SX and SP favor a non-rejection. A plausible explanation for such a difference can be the effect of a small number of influential outliers on the parameter estimates that the J tests rely on, see Vorkink (2003). To investigate whether the results are driven by the outliers we winsorize the excess returns on the 10 portfolios at 0.1%, 0.3%, 0.5%, 0.7%, 0.9%, 1% levels.[8]

Using the entire sample period from January 1965 to December 2009 and different winsorization levels, the results in Table 7 show the sensitivity of the parametric inference to the presence of outliers. For the single-factor model, Panel A, the J tests favor a rejection when the data is winsorized up to 0.1%, and a non-rejection thereafter. In Panel B, the Fama-French three-factor model is rejected by the J_1 and J_4 tests for the low winsorization levels where the inference is reverted as the level increases. For the three-factor case, J_4 appears to be the most sensitive statistic to the outliers, whereas J_2 and J_3 are the least affected ones. The fact that the parametric inference can vary with the presence of influential eccentric observations implies that, one needs to take caution before reaching to a conclusion based on their results.

3.6 Conclusion

The beta-pricing representation of linear factor pricing models is typically assesses with tests based on OLS or GMM. In this context, standard asymptotic theory is known to provide a poor approximation to the finite-sample distribution of those test

[8]We also winsorize the excess returns on the 25 and 100 portfolios. The results are not sensitive to winsorization at any level, hence are not reported here.

statistics, even with fairly large samples. In particular, the asymptotic tests tend to over-reject the null hypothesis when it is in fact true, and these size distortions grow quickly as the number of included test assets increases. So the conclusions of empirical studies that adopt such procedures can be lead to spuriously reject the validity of the asset pricing model.

Exact finite-sample methods that avoid the spurious rejection problem usually rely on strong distributional assumptions about the model's disturbance terms. A prominent example is the GRS test that assumes that the disturbance terms are identically distributed each period according to a multivariate normal distribution. Yet it is known that financial asset returns are non-normal, exhibiting time-varying variances and excess kurtosis. These stylized facts would put into question the reliability of any inference method that assumes that the cross-sectional distribution of disturbance terms is homogenous over time. Another issue with standard inference methods has to do with the choice of how many tests assets to include. Indeed, if too many are included relative to number of available time-series observations, the GRS test may lose all its power or may not even be computable. In fact, any procedure that relies on unrestricted estimates of the covariance matrix of regression disturbances will no longer be computable owing to the singularity that occurs when the size of the cross-section exceeds the length of the time series.

In this paper we have proposed a finite-sample test procedure that overcomes these problems. Specifically, our statistical framework makes no parametric assumption about the distribution of the disturbance terms in the factor model. The only requirement is that the cross-section disturbance vectors be diagonally symmetric each period. The class of diagonally symmetric distributions includes elliptically symmetric ones, which are theoretically consistent with mean-variance analysis. Our nonparametric framework leaves open the possibility of unknown forms of time-varying non-normalities and many other distribution heterogeneities, such as time-varying covariance structures, time-varying kurtosis, etc. The procedure is an adaptive one based on a split-sample technique that is applicable even in large cross-sections. In fact, the power of the new test procedure increases as either the time series lengthens of the cross-section becomes larger. The inference procedure developed here thus adds a potentially very useful way to assess linear factor pricing models.

References

Abdelkhalek, T. and J.-M. Dufour. 1998. “Statistical inference for computable general equilibrium models, with application to a model of the Maroccan economy.” *Review of Economics and Statistics* 80: 520–534.

Affleck-Graves, J. and B. McDonald. 1989. “Nonnormalities and tests of asset pricing theories.” *Journal of Finance* 44: 889–908.

Amsler, C.E. and P. Schmidt. 1985. “A Monte Carlo investigation of the accuracy of multivariate CAPM tests.” *Journal of Financial Economics* 14: 359–375.

Bassett G. and Koenker, R. 1978. “Asymptotic Theory of Least Absolute Error Regression.” of the American Statistical Association 73: 618–622.

Beaulieu, M.-C., J.-M. Dufour, and L. Khalaf. 2007. “Multivariate tests of mean-variance efficiency with possibly non-Gaussian errors.” *Journal of Business and Economic Statistics* 25: 398-410.

Behboodian, J. 1990. “Some characterization theorems on symmetry.” *Computational Statistics and Data Analysis* 10: 189–192.

Berk, J. 1997. “Necessary conditions for the CAPM.” *Journal of Economic Theory* 73: 245–257.

Blattberg, R. and N. Gonedes. 1974. “A comparison of the stable and student distributions as statistical models of stock prices.” *Journal of Business* 47: 244–280.

Boldin, M.V., G.I. Simonova, and Y.N. Tyurin. 1997. Sign-based methods in Linear Statistical Models. American Mathematical Society, Maryland.

Bossaerts, P. and P. Hillion. (1997). "Local parametric analysis of hedging in discrete time." *Journal of Econometrics* 81: 243–272.

Breeden, D.T. 1979. "An intertemporal asset pricing model with stochastic consumption and investment opportunities." *Journal of Financial Economics* 7: 265–296.

Campbell, J.Y., A.W. Lo, and A.C. MacKinlay. 1997. The Econometrics of Financial Markets. Princeton University Press, New Jersey.

Chamberlain, G. 1983. "A characterization of the distributions that imply mean-variance utility functions." *Journal of Economic Theory* 29: 185–201.

Chou, P.-H. and G. Zhou. 2006. "Using bootstrap to test portfolio efficiency." *Anals of Economics and Finance* 2: 217–249.

Connor, G. and R.A. Korajczyk. 2009. "Factor models in portfolio and asset pricing theory." In Handbook of Portfolio Construction: Contemporary Applications of Markowitz Techniques. (J. Guerard, ed.), London: Springer.

Coudin, E. and J.-M. Dufour. 2009. "Finite-sample distribution-free inference in linear median regressions under heteroscedasticity and non-linear dependence of unknown form." *Econometrics Journal* 12: 19–49.

Duffee, G.R. 1995. "Stock returns and volatility: a firm-level analysis." *Journal of Financial Economics* 37: 399–420.

Duffee, G.R. 2001. "Asymmetric cross-sectional dispersion in stock returns: evidence and implications." Federal Reserve Bank of San Francisco Working Paper No. 2000-18.

Dufour, J.-M. 2003. "Identification, weak instruments, and statistical inference in econometrics." *Canadian Journal of Economics* 36: 767–808.

Dufour, J.-M. and M. Hallin. 1991. “Nonuniform bounds for nonparametric t tests.” *Econometric Theory* 7: 253–263.

Dufour, J.-M. and J. Jasiak. 2001. “Finite sample limited information inference methods for structural equations and models with generated regressors.” *International Economic Review* 42: 815–843.

Dufour, J.-M. and L. Khalaf. 2002. “Simulation-based finite- and large-sample tests in multivariate regressions.” *Journal of Econometrics* 111: 303–322.

Dufour, J.-M. and M. Taamouti. 2005. “Projection-based statistical inference in linear structural models with possibly weak instruments.” *Econometrica* 73: 1351–1365.

Dufour, J.-M. and A. Taamouti. 2010. “Exact optimal inference in regression models under heteroskedasticity and non-normality of unknown form.” *Computational Statistics Data Analysis* forthcoming.

Fama, E.F. and K.R. French. 1993. “Common Risk Factors in the Returns on Stocks and Bonds.” *Journal of Financial Economics* 33: 356.

Fama, E. 1965. “The behavior of stock market prices.” *Journal of Business* 38: 34–105.

Ferson, W.E. and S.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.

Gibbons, M.R., S.A. Ross, and J. Shanken. 1989. “A test of the efficiency of a given portfolio.” *Econometrica* 57: 1121–1152.

Glahe, F.R. and J.F. Hunt. 1970. “The small sample properties of simultaneous equation least absolute estimators vis-a-vis least squares estimators.” *Econometrica* 38: 742–753.

Grinblatt, M. and S. Titman. 1987. "The relation between mean-variance efficiency and arbitrage pricing." *Journal of Business* 60: 97–112.

Gungor, S. and R. Luger. 2009. "Exact distribution-free tests of mean-variance efficiency." *Journal of Empirical Finance* 16: 816–829.

Harvey, C.R. and G. Zhou. 1990. "Bayesian inference in asset pricing tests." *Journal of Financial Economics* 26: 221–254.

Hsu, D.A. 1982. "A Bayesian robust detection of shift in the risk structure of stock market returns." *Journal of the American Statistical Association* 77: 29–39.

Huberman, G., S. Kandel, and R.F. Stambaugh. 1987. "Mimicking portfolios and exact arbitrage pricing." *Journal of Finance* 42: 1-9.

Hunt, J.G., J.M. Dowling, and F.R. Glahe. 1974. "L_1 estimation in small samples with Laplace error distributions." *Decision Sciences* 5: 22–29.

Jagannathan, R. and T. Ma. 2003. "Risk reduction in large portfolios: why imposing the wrong constraints helps." *Journal of Finance* 58: 1651–1684.

Jobson, J.D. 1982. "A multivariate linear regression test for the Arbitrage Pricing Theory." *Journal of Finance* 37: 1037–1042 .

Jobson, J.D. and B. Korkie. 1982. "Potential performance and tests of portfolio efficiency." *Journal of Financial Economics* 10: 433–466.

Jobson, J.D. and B. Korkie. 1985. "Some tests of linear asset pricing with multivariate normality." *Canadian Journal of Administrative Sciences* 2: 114–138.

Kandel, S., R. McCulloch, and R.F. Stambaugh. 1995. "Bayesian inference and portfolio efficiency." *Review of Financial Studies* 8: 1–53.

Kocherlakota, N.R. 1997. " Testing the Consumption CAPM with heavy-tailed pricing errors." *Macroeconomic Dynamics* 1: 551-567.

Koenker, R. and Bassett G. 1978. "Regression Quantiles." *Econometrica* 46: 33–50.

Ledoit, O. and M. Wolf. 2003. "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection." *Journal of Empirical Finance* 10: 603–621.

Ledoit, O. and M. Wolf. 2004. "Honey, I shrunk the sample covariance matrix." *Journal of Portfolio Management* 30: 110–119.

Lehmann, E.L. and J.P. Romano. 2005. Testing Statistical Hypotheses, Third Edition. Springer, New York.

Lehmann, E.L. and C. Stein. 1949. "On the theory of some non-parametric hypotheses." *Annals of Mathematical Statisitcs* 20: 28–45.

Lintner, J. 1965. "The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets." *Review of Economics and Statistics* 47: 13–37.

Lo, A. and A.C. MacKinlay. 1990. "Data-snooping biases in tests of financial asset pricing models." *Review of Financial Studies* 3: 431–467.

MacKinlay, A.C. 1987. "On multivariate tests of the Capital Asset Pricing Model." *Journal of Financial Economics* 18: 341–372.

MacKinlay, A.C. and M.P. Richardson. 1991. "Using Generalized Method of Moments to test mean-variance efficiency." *Journal of Finance* 46: 511–527.

Merton, R.C. 1973. "An Intertemporal Capital Asset Pricing Model." *Econometrica* 41: 867–887.

Michaud, R. 1989. "The Markowitz optimization enigma: is 'optimized' optimal?" *Financial Analysts Journal* 45: 31–42.

Mitra, A. 1987. "Minimum absolute deviation estimation in regression analysis." *Computers and Industrial Engineering* 12: 67–76.

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

Owen, J. and R. Rabinovitch. 1983. "On the class of elliptical distributions and their applications to the theory of portfolio choice." *Journal of Finance* 38: 745–752.

Pfaffenberger, R.C. and J.J. Dinkel. 1978. "Absolute deviations curve fitting: an alternative to least squares." In Contributions to Survey Sampling and Applied Statistics (H.A. David, ed.) Academic Press, New York.

Pratt, J.W. and J.D. Gibbons. 1981. Concepts of Nonparametric Theory. Springer, New York.

Randles, R.H. and D.A. Wolfe. 1979. Introduction to the Theory of Nonparametric Statistics. Wiley, New York.

Roll, R. 1979. "A reply to Mayers and Rice (1979)." *Journal of Financial Economics* 7: 391–400.

Rosenberg, B. and D. Carlson. 1977. "A simple approximation of the sampling distribution of least absolute residuals regression estimates." *Communications in Statistics - Simulation and Computation* 6: 421–437.

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

Sentana, E. 2009. "The econometrics of mean-variance efficiency tests: a survey." *Econometrics Journal* 12: 65–101.

Serfling, R.J. 2006. "Multivariate symmetry and asymmetry." In Encyclopedia of Statistical Sciences, Second Edition (S. Kotz, N. Balakrishnan, C. B. Read and B. Vidakovic, eds.), Wiley.

Shanken, J. 1987. "A Bayesian approach to testing portfolio efficiency." *Journal of Financial Economics* 19: 195–216.

Shanken, J. 1996. "Statistical methods in tests of portfolio efficiency: a synthesis." In Handbook of Statistics, Vol. 14: Statistical Methods in Finance. (G.S. Maddala and C.R. Rao, eds.), North-Holland, Amsterdam.

Sharpe, W.F. 1964. "Capital asset prices: a theory of market equilibrium under conditions of risk." *Journal of Finance* 19: 425–442.

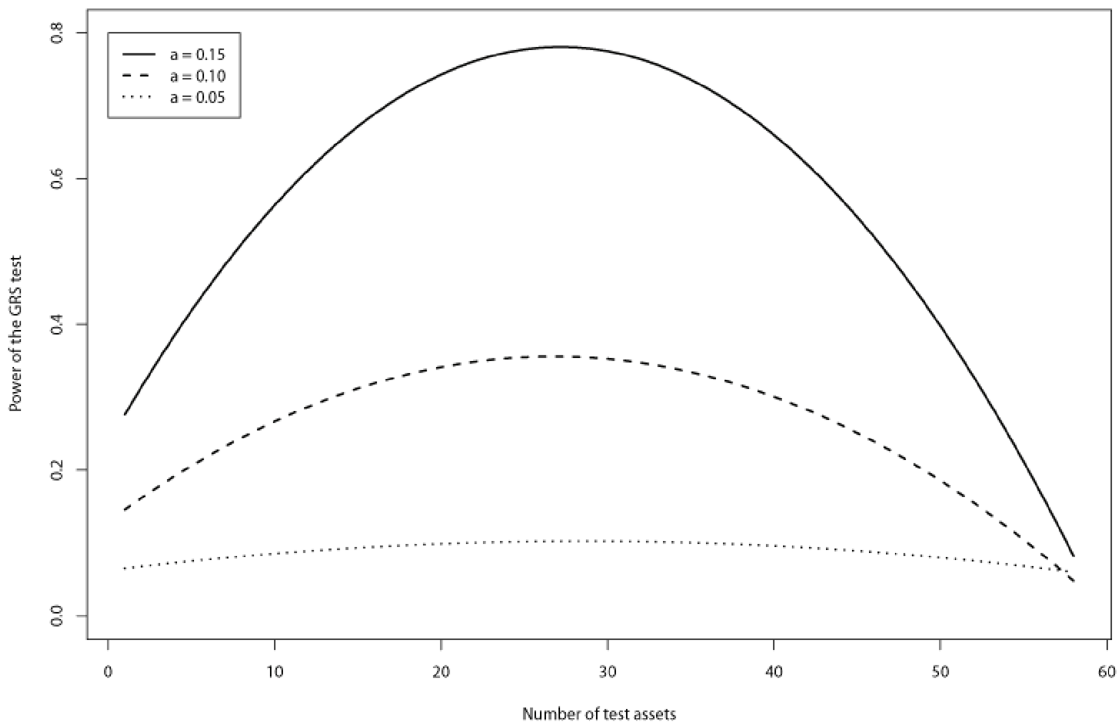
Stambaugh, R.F. 1982. "On the exclusion of assets from tests of the two-parameter model: a sensitivity analysis." *Journal of Financial Economics* 10: 237–268.

Stewart, K.G. 1997. "Exact testing in multivariate regression." *Econometric Reviews* 16: 321–352.

White, H. 1980. "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity ." *Econometrica* 48: 817–838.

Zhou, G. 1993. "Asset-pricing tests under alternative distributions." *Journal of Finance* 48: 1927–1942.

Figure 3.1: **Power of the GRS Test.**



Notes: This figure plots the power of the GRS test as a function of the number of included test assets. The returns are generated from a single-factor model with normally distributed disturbances. The sample size is $T = 60$ and the number of test assets N ranges from 1 to 58. The test is performed at a nominal 0.05 level. The higher power curves are associated with greater pricing errors.

Table 3.1: **Empirical power comparisons for various sample splits.**

T_1/T	0.2	0.3	0.4	0.5	0.6	0.7	0.8
Panel A: t(12) Distribution							
$\|a_i\| = 0.20$							
SX	85.9	91.9	89.2	84.7	76.3	50.7	10.9
SP	95.5	97.7	98.3	97.6	97.1	81.7	34.1
$\|a_i\| = 0.15$							
SX	37.5	46.9	49.3	43.3	35.7	19.1	3.2
SP	56.9	69.8	67.9	65.8	60.1	39.1	12.1
$\|a_i\| = 0.10$							
SX	6.6	7.2	9.2	7.8	7.1	3.2	0.8
SP	12.5	16.9	16.5	14.9	14.8	7.7	2.2
Panel B: t(6) Distribution							
$\|a_i\| = 0.20$							
SX	71.5	81.9	81.4	76.0	62.8	36.6	7.5
SP	87.1	93.5	95.5	94.4	86.9	68.3	25.4
$\|a_i\| = 0.15$							
SX	24.7	32.6	32.9	31.1	25.3	12.0	1.9
SP	41.4	53.7	52.3	54.8	44.1	27.5	9.6
$\|a_i\| = 0.10$							
SX	4.0	6.2	4.5	4.2	4.5	2.2	0.3
SP	8.1	12.7	9.6	8.9	8.7	5.1	1.6

Notes: This table reports the empirical power (in percentages) of the proposed test procedure based on the SX and SP statistics in (3.3.5) and (3.3.6) for various sample splits, T_1/T. The sample size is $T = 60$ and the number of test assets is $N = 100$. The returns are generated according to a single-factor model in with i.i.d. disturbances following a Student-t distribution with degrees of freedom equal to 12 (Panel A) or 6 (Panel B). The notation $|a_i| = a$ means that $N/2$ pricing errors are set as $a_i = -a$ and the other half are set as $a_i = a$. The nominal level is 0.05 and the results are based on 1000 replications.

Table 3.2: **Empirical size and power comparisons with homoskedastic disturbances.**

T	N	J_1	J_2	J_3	J_4	SD	WD	SX	SP
Panel A: Size									
60	10	4.6	9.5	4.7	7.9	5.5	4.2	0.2	0.9
	25	4.3	32.1	5.6	14.9	4.4	5.2	0.4	1.4
	50	6.0	98.7	41.2	-	5.2	4.1	0.7	1.1
	100	-	-	-	-	4.6	2.9	0.5	0.9
	125	-	-	-	-	4.5	4.2	1.0	2.0
120	10	3.8	5.8	3.8	5.2	3.8	4.2	0.1	1.2
	25	5.1	12.0	5.2	7.8	4.2	3.4	0.5	1.0
	50	5.5	45.0	7.5	13.9	5.2	4.8	0.6	1.5
	100	4.0	100.0	53.5	-	4.2	3.7	0.8	1.8
	125	-	-	-	-	6.1	5.0	0.7	1.7
Panel B: Size-corrected power									
60	10	41.5	41.5	41.5	41.9	16.7	18.4	1.7	4.5
	25	55.5	55.5	55.5	52.7	26.0	25.3	5.7	11.0
	50	24.1	24.1	24.1	-	40.8	43.4	14.2	29.3
	100	-	-	-	-	61.6	69.0	37.2	55.9
	125	-	-	-	-	67.3	75.2	44.5	64.0
120	10	83.1	83.1	83.1	83.8	32.6	39.0	11.5	22.1
	25	97.0	97.0	97.0	97.4	55.9	62.4	33.2	51.4
	50	99.6	99.6	99.6	99.5	79.4	86.6	66.2	83.4
	100	91.1	91.1	91.1	-	96.4	98.2	94.4	98.7
	125	-	-	-	-	98.7	99.1	98.4	99.7

Notes: This table reports the empirical size (Panel A) and size-corrected power (Panel B) of the GRS test (J_1), the LR test (J_2), an adjusted LR test (J_3), a GMM-based test (J_4), a sign test (SD), a Wilcoxon signed rank test (SD), and the proposed SX- and SP-based tests. The returns are generated according to a single-factor model with i.i.d. disturbances following a $t(6)$ distribution. The pricing errors are zero under H_0, whereas $N/2$ pricing errors are set equal to -0.15 and the other half are set to 0.15 under H_1. The nominal level is 0.05 and entries are percentage rates. The results are based on 1000 replications and the symbol "-" is used whenever a test is not computable.

Table 3.3: **Empirical size and power comparisons with contemporaneous heteoskedastic disturbances.**

T	N	J_1	J_2	J_3	J_4	SD	WD	SX	SP
Panel A: Size									
60	10	23.2	32.8	23.6	26.6	5.9	5.3	0.4	0.9
	25	46.6	81.6	50.5	62.8	5.5	4.3	0.3	1.1
	50	50.6	97.4	90.2	-	3.9	4.4	0.9	2.1
	100	-	-	-	-	4.5	3.7	1.2	2.3
	200	-	-	-	-	5.4	4.2	1.1	2.6
120	10	19.0	23.1	19.1	18.5	4.9	4.4	0.3	1.5
	25	37.8	54.6	38.3	45.5	4.2	5.0	0.6	1.8
	50	67.8	92.8	72.2	78.2	4.8	3.9	1.1	2.4
	100	73.7	96.6	94.0	-	5.7	5.2	1.8	2.2
	200	-	-	-	-	6.0	4.9	1.6	1.9
Panel B: Size-corrected power									
60	10	14.9	14.9	14.9	15.0	14.0	15.9	2.0	4.2
	25	27.2	27.2	27.2	25.1	20.6	26.0	4.0	6.3
	50	32.3	32.3	32.3	-	28.6	35.4	6.7	11.6
	100	-	-	-	-	49.4	59.3	15.4	21.7
	200	-	-	-	-	72.4	80.3	29.9	38.8
120	10	23.0	23.0	23.0	24.0	26.8	31.2	7.3	10.8
	25	47.8	47.8	47.8	45.7	43.7	51.7	17.6	23.2
	50	78.6	78.6	78.6	73.6	69.4	78.9	36.0	44.2
	100	76.3	76.3	76.3	-	91.0	95.7	63.9	71.2
	200	-	-	-	-	99.6	99.8	88.6	91.9

Notes: This table reports the empirical size (Panel A) and size-corrected power (Panel B) of the GRS test (J_1), the LR test (J_2), an adjusted LR test (J_3), a GMM-based test (J_4), a sign test (SD), a Wilcoxon signed rank test (SD), and the proposed SX- and SP-based tests. The returns are generated according to a single-factor model with contemporaneous heteroskedastic disturbances. The pricing errors are zero under H_0, whereas $N/2$ pricing errors are set equal to -0.15 and the other half are set to 0.15 under H_1. The nominal level is 0.05 and entries are percentage rates. The results are based on 1000 replications and the symbol "-" is used whenever a test is not computable.

Table 3.4: **Empirical size under cross-sectional disturbance equicorrelation structure.**

	Panel A: $N = 10$					Panel B: $N = 100$				
ρ	0.1	0.2	0.3	0.4	0.5	0.1	0.2	0.3	0.4	0.5
$T = 60$										
SD	4.5	5.2	5.7	7.1	6.9	5.5	10.1	12.3	15.4	17.7
WD	5.3	5.3	6.1	8.2	8.8	6.6	12.0	15.2	16.9	19.8
SX	0.7	0.1	0.1	0.1	0.1	0.2	0.3	0.1	0.4	0.0
SP	1.3	0.5	0.6	0.5	0.4	0.9	1.1	0.5	1.2	0.3
$T = 120$										
SD	4.9	6.5	6.8	7.7	8.9	8.7	13.5	17.6	19.0	20.6
WD	4.8	6.2	6.6	7.8	10.0	10.1	13.4	17.0	19.5	21.9
SX	0.0	0.1	0.2	0.3	0.5	0.7	0.6	0.3	0.4	0.6
SP	0.3	0.7	0.9	0.9	1.2	1.9	1.1	0.7	0.8	1.0

Notes: This table reports the empirical size of a sign test (SD), a Wilcoxon signed rank test (SD), and the proposed SX- and SP-based tests when the cross-sectional disturbances are multivariate normal with mean zero and the correlation between any two disturbances is equal to ρ. The nominal level is 0.05 and entries are percentage rates. The results are based on 1000 replications

Table 3.5: **Test results for the single-factor model with 10 portfolios formed on size.**

Time	J_1	J_2	J_3	J_4	SD	WD	SX	SP
45-year period								
1/65–12/09	2.30 (0.01)	22.94 (0.01)	22.64 (0.01)	23.04 (0.01)	5.91 (0.82)	8.31 (0.60)	[-1.34, 1.44]	[-0.40, 0.16]
5-year subperiods								
1/65–12/69	1.80 (0.08)	18.80 (0.04)	16.61 (0.08)	18.42 (0.05)	24.00 (0.01)	34.80 (0.00)	[-4.79, 5.20]	[-0.48, 1.37]
1/70–12/74	1.45 (0.19)	15.52 (0.11)	13.71 (0.19)	14.27 (0.16)	34.93 (0.00)	34.94 (0.00)	[-3.72, 6.27]	[-0.25, 2.30]
1/75–12/79	1.30 (0.26)	14.09 (0.17)	12.45 (0.26)	18.50 (0.05)	33.07 (0.00)	58.75 (0.00)	[-3.86, 6.13]	[0.28, 1.72]
1/80–12/84	1.17 (0.33)	12.84 (0.23)	11.34 (0.33)	11.99 (0.29)	3.07 (0.98)	5.79 (0.83)	[-4.87, 5.12]	[-0.45, 0.78]
1/85–12/89	1.27 (0.27)	13.85 (0.18)	12.23 (0.27)	12.60 (0.25)	9.60 (0.48)	4.33 (0.93)	[-4.73, 5.26]	[-0.31, 1.18]
1/90–12/94	1.80 (0.08)	18.80 (0.04)	16.61 (0.08)	18.28 (0.05)	8.93 (0.54)	14.01 (0.17)	[-1.45, 0.66]	[-0.24, 0.45]
1/95–12/99	1.60 (0.14)	16.94 (0.08)	14.96 (0.13)	20.20 (0.03)	9.47 (0.49)	7.29 (0.70)	[-3.75, 6.24]	[-1.10, 2.09]
1/00–12/04	1.93 (0.06)	19.93 (0.03)	17.61 (0.06)	19.46 (0.03)	27.60 (0.00)	16.80 (0.08)	[-4.66, 5.33]	[-0.88, 2.18]
1/05–12/09	1.54 (0.15)	16.39 (0.09)	14.48 (0.15)	15.42 (0.12)	8.93 (0.54)	6.57 (0.77)	[-5.07, 4.92]	[-1.07, 1.33]
10-year subperiods								
1/65–12/74	1.26 (0.26)	13.17 (0.21)	12.40 (0.26)	12.50 (0.25)	6.73 (0.75)	3.34 (0.97)	[-5.80, 4.19]	[-1.89, 0.05]
1/75–12/84	1.81 (0.07)	18.45 (0.05)	17.37 (0.07)	23.18 (0.01)	40.87 (0.00)	56.62 (0.00)	[-4.69, 5.30]	[-0.19, 0.85]
1/85–12/94	2.29 (0.02)	22.84 (0.01)	21.511 (0.02)	22.18 (0.01)	7.20 (0.71)	3.85 (0.95)	[-1.07, -0.21]	[-1.06, -0.54]
1/95–12/04	2.30 (0.02)	22.94 (0.01)	21.60 (0.02)	21.61 (0.02)	10.00 (0.44)	7.61 (0.67)	[-6.11, 3.88]	[-2.29, 0.66]

Notes: The results are based on value-weighted returns of 10 portfolios formed on size. The market portfolio is the value-weighted return on all NYSE, AMEX, and NASDAQ stocks and the risk-free rate is the one-month Treasury bill rate. Columns 2–5 report the results for the parametric J test; columns 6–7 report the results for the non-parametric SD and WD statistics. The numbers in parantheses are the p-values. The results for the newly proposed procedure, SX and SP, are reported in columns 8 and 9. The 95% marginal confidence intervals of the intercept estimates are in square brackets.

Table 3.6: **Test results for the Fama-French model with 10 portfolios formed on size.**

Time	J_1	J_2	J_3	J_4	SX	SP
45-year period						
1/65–12/09	2.44 (0.01)	24.43 (0.01)	24.02 (0.01)	24.99 (0.01)	[-0.59, 0.69]	[-0.03, 0.21]
5-year subperiods						
1/65–12/69	1.03 (0.44)	11.87 (0.29)	10.09 (0.43)	12.18 (0.27)	[-1.00, 0.96]	[-0.48, 0.40]
1/70–12/74	1.39 (0.21)	15.59 (0.11)	13.25 (0.21)	19.55 (0.03)	[-0.88, 1.08]	[-0.32, 0.44]
1/75–12/79	0.41 (0.93)	5.08 (0.89)	4.31 (0.93)	6.09 (0.81)	[-1.04, 0.92]	[-0.60, 0.48]
1/80–12/84	1.92 (0.07)	20.55 (0.02)	17.47 (0.06)	28.76 (0.00)	[-0.71, 1.29]	[-0.35, 0.61]
1/85–12/89	1.51 (0.16)	16.75 (0.08)	14.24 (0.16)	19.07 (0.04)	[-0.73, 1.27]	[-0.13, 0.67]
1/90–12/94	1.70 (0.11)	18.56 (0.05)	15.78 (0.11)	21.81 (0.02)	[-0.86, 1.10]	[-0.34, 0.54]
1/95–12/99	1.70 (0.11)	18.50 (0.05)	15.72 (0.11)	31.27 (0.00)	[-0.70, 1.30]	[-0.14, 0.66]
1/00–12/04	1.29 (0.26)	14.58 (0.15)	12.39 (0.26)	17.72 (0.06)	[-1.15, 0.81]	[-0.91, 0.81]
1/05–12/09	1.52 (0.16)	16.82 (0.08)	14.29 (0.16)	19.54 (0.03)	[-1.16, 0.80]	[-0.44, 0.20]
10-year subperiods						
1/65–12/74	1.06 (0.40)	11.35 (0.33)	10.50 (0.40)	12.52 (0.25)	[-1.04, 0.92]	[-0.16, 0.20]
1/75–12/84	0.95 (0.50)	10.16 (0.43)	9.39 (0.50)	12.28 (0.27)	[-1.13, 0.83]	[-0.45, 0.19]
1/85–12/94	2.56 (0.01)	25.71 (0.00)	23.79 (0.01)	26.66 (0.00)	[-0.94, 1.02]	[-0.18, 0.34]
1/95–12/04	2.16 (0.03)	22.07 (0.01)	20.42 (0.03)	25.30 (0.00)	[-1.18, 0.78]	[-0.54, 0.30]

Note: The results are based on value-weighted returns of 10 portfolios formed on size, the returns on three Fama-French factors, and the one-month Treasury bill rate as the risk-free rate. Columns 2–5 report the results for the parametric J statistics and the p-values in the paranthesis. The results for the newly proposed distribution-free tests, SX and SP, are reported in columns 6 and 7. The 95% marginal confidence intervals of the intercept estimates are in square brackets.

Table 3.7: **Test results for the single-factor model and 25 portfolios formed on size and book-to-market.**

Time	J_1	J_2	J_3	J_4	SD	WD	SX	SP
45-year period								
1/65–12/09	4.18 (0.00)	99.93 (0.00)	97.25 (0.00)	106.49 (0.00)	97.13 (0.00)	121.42 (0.00)	[-1.14, 1.48]	[-0.21, 0.27]
5-year subperiods								
1/65–12/69	1.90 (0.04)	52.53 (0.00)	39.83 (0.03)	46.81 (0.01)	34.53 (0.10)	58.61 (0.00)	[-4.80, 5.19]	[-0.50, 1.39]
1/70–12/74	1.56 (0.11)	45.84 (0.01)	34.76 (0.09)	38.07 (0.05)	42.00 (0.02)	42.53 (0.02)	[-3.63, 6.36]	[-0.30, 2.41]
1/75–12/79	0.88 (0.62)	30.03 (0.22)	22.77 (0.59)	28.01 (0.31)	65.87 (0.00)	116.08 (0.00)	[-4.53, 5.46]	[-0.22, 1.00]
1/80–12/84	1.73 (0.07)	49.17 (0.00)	37.29 (0.05)	44.33 (0.01)	21.87 (0.64)	23.02 (0.58)	[-4.54, 5.45]	[-0.13, 1.06]
1/85–12/89	2.51 (0.01)	62.68 (0.00)	47.53 (0.00)	77.33 (0.00)	66.67 (0.00)	67.67 (0.00)	[-4.61, 5.38]	[-0.03, 0.96]
1/90–12/94	2.03 (0.03)	54.74 (0.00)	41.51 (0.02)	51.85 (0.00)	38.80 (0.04)	73.94 (0.00)	[-3.22, 1.11]	[-0.88, 0.40]
1/95–12/99	1.20 (0.31)	37.92 (0.05)	28.76 (0.27)	42.34 (0.02)	13.87 (0.96)	19.87 (0.75)	[-4.46, 5.53]	[-0.73, 1.36]
1/00–12/04	1.57 (0.11)	46.10 (0.01)	34.96 (0.09)	39.80 (0.03)	87.33 (0.00)	49.27 (0.00)	[-4.40, 5.59]	[-0.31, 1.74]
1/05–12/09	1.60 (0.10)	46.68 (0.01)	35.40 (0.08)	40.05 (0.03)	24.27 (0.50)	19.27 (0.78)	[-5.05, 4.94]	[-0.73, 0.40]
10-year subperiods								
1/65–12/74	1.62 (0.05)	43.03 (0.01)	37.83 (0.05)	39.87 (0.03)	31.40 (0.18)	29.06 (0.26)	[-5.74, 4.25]	[-1.61, 0.03]
1/75–12/84	1.34 (0.16)	36.54 (0.06)	32.13 (0.15)	39.56 (0.03)	86.60 (0.00)	116.16 (0.00)	[-4.50, 5.49]	[-0.02, 1.02]
1/85–12/94	4.41 (0.00)	93.08 (0.00)	81.84 (0.00)	116.37 (0.00)	74.80 (0.00)	84.90 (0.00)	[-2.00, 1.89]	[-0.19, 0.65]
1/95–12/04	2.59 (0.00)	62.88 (0.00)	55.28 (0.00)	61.81 (0.00)	57.40 (0.00)	62.42 (0.00)	[-5.41, 4.58]	[-1.37, 1.00]

Note: The results are based on value-weighted returns of 25 portfolios formed on size and book-to-market. The market portfolio is the value-weighted return on all NYSE, AMEX, and NASDAQ stocks and the risk-free rate is the one-month Treasury bill rate. Columns 2–5 report the results for the parametric J test; columns 6–7 report the results for the non-parametric SD and WD statistics. The numbers in parantheses are the p-values. The results for the newly proposed procedure, SX and SP, are reported in columns 8 and 9. The 95% marginal confidence intervals of the intercept estimates are in square brackets.

Table 3.8: **Test results for the Fama-French Model with 25 portfolios formed on size and book-to-market.**

Time	J_1	J_2	J_3	J_4	SX	SP
45-year period						
1/65–12/09	3.26 (0.00)	79.71 (0.00)	77.27 (0.00)	87.00 (0.00)	[-0.28, 0.32]	[0.08, 0.16]
5-year subperiods						
1/65–12/69	1.44 (0.16)	45.20 (0.01)	32.77 (0.14)	-	[-1.04, 0.92]	[-0.36, 0.48]
1/70–12/74	1.48 (0.15)	46.03 (0.01)	33.37 (0.12)	-	[-0.90, 1.06]	[-0.38, 0.54]
1/75–12/79	0.58 (0.92)	22.31 (0.62)	16.18 (0.91)	-	[-1.16, 0.80]	[-0.48, 0.56]
1/80–12/84	1.27 (0.26)	41.31 (0.02)	29.95 (0.23)	-	[-0.78, 1.22]	[-0.22, 0.62]
1/85–12/89	2.18 (0.02)	59.61 (0.00)	43.22 (0.01)	-	[-0.85, 1.11]	[-0.05, 0.55]
1/90–12/94	2.25 (0.02)	60.93 (0.00)	44.18 (0.01)	-	[-0.88, 1.08]	[-0.36, 0.56]
1/95–12/99	1.12 (0.38)	37.71 (0.05)	27.34 (0.34)	-	[-0.95, 1.01]	[-0.11, 0.61]
1/00–12/04	1.11 (0.38)	37.54 (0.05)	27.21 (0.35)	-	[-0.92, 1.04]	[-0.48, 0.48]
1/05–12/09	1.56 (0.12)	47.73 (0.00)	34.61 (0.10)	-	[-0.99, 0.97]	[-0.23, 0.37]
10-year subperiods						
1/65–12/74	1.42 (0.12)	39.13 (0.04)	33.75 (0.11)	48.74 (0.00)	[-0.99, 0.97]	[-0.07, 0.13]
1/75–12/84	0.78 (0.76)	22.98 (0.58)	19.82 (0.76)	28.19 (0.30)	[-0.99, 0.97]	[-0.19, 0.09]
1/85–12/94	4.04 (0.00)	88.95 (0.00)	76.72 (0.00)	139.18 (0.00)	[-0.80, 1.16]	[-0.08, 0.56]
1/95–12/04	2.09 (0.01)	53.95 (0.00)	46.53 (0.01)	88.08 (0.00)	[-0.95, 1.01]	[-0.27, 0.49]

Note: The results are based on value-weighted returns of 25 portfolios formed on size and book-to-market, the returns on three Fama-French factors, and the one-month Treasury bill rate as the risk-free rate. Columns 2–5 report the results for the parametric J statistics and the p-values in the paranthesis. The results for the newly proposed distribution-free tests, SX and SP, are reported in columns 6 and 7. The 95% marginal confidence intervals of the intercept estimates are in square brackets.

Table 3.9: **Test results for the single-factor model and 100 portfolios formed on size and book-to-market.**

Time	J_1	J_2	J_3	J_4	SD	WD	SX	SP
45-year period								
1/65–12/09	2.74 (0.00)	262.02 (0.00)	236.79 (0.00)	277.65 (0.00)	260.95 (0.00)	314.09 (0.00)	[-0.92, 1.38]	[-0.12, 0.35]
5-year subperiods								
1/65–12/69	-	-	-	-	143.33 (0.00)	186.92 (0.00)	[-4.68, 5.31]	[-0.35, 1.36]
1/70–12/74	-	-	-	-	152.80 (0.00)	148.81 (0.00)	[-3.90, 6.09]	[-0.36, 1.56]
1/75–12/79	-	-	-	-	227.47 (0.00)	355.90 (0.00)	[-3.03, 6.96]	[0.33, 6.96]
1/80–1/84	-	-	-	-	96.40 (0.58)	96.91 (0.57)	[-4.59, 5.40]	[0.00, 0.71]
1/85–12/89	-	-	-	-	164.00 (0.00)	156.98 (0.00)	[-4.58, 5.41]	[-0.09, 0.68]
1/90–12/94	-	-	-	-	156.67 (0.00)	185.98 (0.00)	[-3.88, 0.83]	[-0.85, 0.25]
1/95–12/99	-	-	-	-	70.93 (0.99)	85.86 (0.84)	[-4.30, 5.69]	[-1.00, 1.57]
1/00–12/04	-	-	-	-	219.47 (0.00)	201.90 (0.00)	[-4.16, 5.83]	[-0.19, 1.47]
1/05–12/09	-	-	-	-	89.07 (0.78)	80.24 (0.93)	[-5.09, 4.90]	[-0.50, 0.50]
10-year subperiods								
1/65–12/74	1.33 (0.24)	249.48 (0.00)	141.37 (0.00)	-	121.27 (0.07)	104.95 (0.35)	[-5.59, 4.40]	[-1.43, 0.09]
1/75–12/84	0.94 (0.60)	214.31 (0.00)	121.44 (0.07)	-	263.00 (0.00)	345.82 (0.00)	[-4.56, 5.43]	[-0.83, 1.06]
1/85–12/94	1.75 (0.08)	278.58 (0.00)	157.86 (0.00)	-	193.20 (0.00)	194.08 (0.00)	[-1.58, 1.78]	[-0.12, 0.60]
1/95–12/04	2.57 (0.01)	321.22 (0.00)	182.03 (0.00)	-	186.07 (0.00)	212.82 (0.00)	[-5.39, 4.60]	[-1.47, 0.53]

Note: The results are based on value-weighted returns of 100 portfolios formed on size and book-to-market. The market portfolio is the value-weighted return on all NYSE, AMEX, and NASDAQ stocks and the risk-free rate is the one-month Treasury bill rate. Columns 2–5 report the results for the parametric J test; columns 6–7 report the results for the distribution-free tests of SD and WD. The number in parantheses are the p-values. The results for the newly proposed procedure, SX and SP, are reported in columns 8 and 9. The 95% marginal confidence intervals of the intercept estimates are in square brackets.

Table 3.10: **Test results for the Fama-French model with 100 portfolios sorted by size and book-to-market.**

Time	J_1	J_2	J_3	J_4	SX	SP
45-year period						
1/65–12/09	2.48 (0.00)	243.09 (0.00)	218.78 (0.00)	327.08 (0.00)	[-0.14, 0.74]	[0.06, 0.22]
5-year subperiods						
1/65–12/69	-	-	-	-	[-0.88, 1.08]	[-0.16, 0.48]
1/70–12/74	-	-	-	-	[-1.15, 0.81]	[-0.35, 0.29]
1/75–12/79	-	-	-	-	[-0.93, 5.07]	[-0.93, 5.08]
1/80–12/84	-	-	-	-	[-1.00, 0.96]	[-0.32, 0.40]
1/85–12/89	-	-	-	-	[-0.94, 1.02]	[-0.22, 0.58]
1/90–12/94	-	-	-	-	[-1.03, 0.93]	[-0.51, 0.29]
1/95–12/99	-	-	-	-	[-0.63, 1.37]	[-0.15, 1.02]
1/00–12/04	-	-	-	-	[-1.37, 2.63]	[-1.37, 2.31]
1/05–12/09	-	-	-	-	[-1.01, 0.95]	[-0.45, 0.63]
10-year subperiods						
1/65–12/74	1.78 (0.09)	297.18 (0.00)	163.45 (0.00)	-	[-0.95, 1.01]	[-0.07, 0.25]
1/75–12/84	0.79 (0.76)	191.69 (0.00)	105.43 (0.34)	-	[-1.92, 2.04]	[-1.08, 1.68]
1/85–12/94	1.67 (0.11)	285.83 (0.00)	157.21 (0.00)	-	[-0.85, 1.11]	[-0.05, 0.39]
1/95–12/04	2.23 (0.03)	317.49 (0.00)	174.62 (0.00)	-	[-0.84, 1.12]	[-0.40, 0.56]

Note: The results are based on value-weighted returns of 100 portfolios formed on size and book-to-market, the returns on three Fama-French factors, and the one-month Treasury bill rate as the risk-free rate. Columns 2–5 report the results for the parametric J statistics and the p-values in the paranthesis. The results for the newly proposed distribution-free tests, SX and SP, are reported in columns 6 and 7. The 95% marginal confidence intervals of the intercept estimates are in square brackets.

Table 3.11: **Sensitivity of parametric tests to extreme observations.**

	0%	0.1%	0.3%	0.5%	0.7%	0.9%	1%
Panel A: Single-factor model							
$J1$	2.30	2.18	1.73	1.61	1.52	1.49	1.48
	(0.01)	(0.02)	(0.07)	(0.10)	(0.13)	(0.14)	(0.14)
$J2$	22.94	21.77	17.41	16.21	15.34	15.03	14.94
	(0.01)	(0.02)	(0.07)	(0.09)	(0.12)	(0.13)	(0.13)
$J3$	22.64	21.49	17.18	16.00	15.14	14.84	14.75
	(0.01)	(0.02)	(0.07)	(0.10)	(0.13)	(0.14)	(0.14)
$J4$	23.04	20.00	14.86	13.67	12.98	12.82	12.73
	(0.01)	(0.03)	(0.14)	(0.19)	(0.22)	(0.23)	(0.24)
Panel B: Fama-French three-factor model							
$J1$	2.39	2.39	1.94	1.77	1.63	1.50	1.46
	(0.01)	(0.01)	(0.04)	(0.06)	(0.09)	(0.13)	(0.15)
$J2$	23.90	23.23	19.39	19.99	20.47	19.02	19.07
	(0.01)	(0.01)	(0.04)	(0.03)	(0.03)	(0.04)	(0.04)
$J3$	23.57	22.85	19.07	19.66	20.13	18.71	18.76
	(0.01)	(0.01)	(0.04)	(0.03)	(0.03)	(0.04)	(0.04)
$J4$	23.92	24.01	18.01	15.36	13.46	12.19	11.73
	(0.01)	(0.01)	(0.05)	(0.12)	(0.20)	(0.27)	(0.30)

Note: The above parametric statistics are based on winsorized excess returns on 10 portfolios at 0.1%, 0.3%, 0.5%, 0.7%, 0.9%, and 1% levels, for the full sample period from Janury 1965 to December 2009. The p-values associated with each test statistic are reported in parantheses.