Personal Asset Pricing and the Premium Investment Framework

Filip Bekjarovski *

February 2018

Abstract

Rational anomaly explanation should be tested out-of-sample. The importance of the approach is demonstrated by showing that isolated risk testing has led to inaccurate conclusions regarding the importance of recession and crash risk for cross-sectional stock anomalies. Isolated risk testing is grounded in a popular misconception that passive and active bets are substitutes. Adopting the complements interpretation also challenges results in comparison studies of factors and traditional investing as well as findings in investigations that rely on excess returns rather than alphas. The paper develops personal asset pricing to motivate the use of benchmarks when the overlapping equilibrium model is not testable and derives the conditions under which anomalies are a free lunch in a frictionless world.

Keywords: anomaly evaluation, recession risk, crash risk, alpha decay, investor heterogeneity.

JEL Classification: G10, G11, G12.

* Amundi, Tilburg University, and University Toulouse 1 Capitole.
Email: f.bekjarovski@tilburguniversity.edu.
Address: Tilburg University, PO Box 90153 5000 LE Tilburg;

1. Introduction

Rational anomaly explanations are frequently tested by regressing the performance of a single anomaly on a new 'rational benchmark'. Combined with risk model mining, this form of isolated risk testing has led to erroneous conclusions regarding the relevance of risk models. The paper proposes assessing rational models on alternative anomalies or portfolios of premiums as an out-of-sample test and empirically demonstrates the importance of the approach. For example, when it comes to (cross-sectional stock) momentum, it is often argued that it is subject to extreme crashes (Barroso & Santa-Clara, 2015; Daniel & Moskowitz, 2016) or that it performs poorly in recession states (Chordia & Shivakumar, 2002). Using the out-of-sample approach shows that crashes and recessions are anomaly specific and not valid risk explanations for cross-sectional stock anomalies. Multi-premium portfolios manage to curtail crash risk to passive investing levels and to significantly *improve* recession performance. The empirical investigation demonstrates the importance of out-of-sample testing for assessing new rational explanations. Moreover, it shows the effectiveness of portfolios as a simple universal tool for reducing idiosyncratic premium risks.

The paper argues that the frequent use of isolated risk testing arises due to an often-held belief that anomalies are substitutes rather than complements to passive investments. A belief that has also motivated comparison investigations between factors and traditional investing. Even more problematically, it has promoted widespread use of excess returns rather than alphas as an anomaly assessment metric in the literature. The paper shows that well known findings in the literature no longer hold under the complements interpretation which is more consistent with the purpose and historical development of anomalies. For example, anomaly alphas do not show strong signs of decay. This is contrary to evidence of excess returns decay in McLean and Pontiff (2016). Alternatively, the results show that passive-active premium combinations are superior to passive investing across all performance assessment dimensions. This is contrary to findings in Idzorek and Kowara (2013) who compare active to passive investing to obtain inconclusive evidence. The objective of the empirical investigation is to show how the substitutes interpretation has motivated the use of imperfect methods in the literature that meaningfully alter analysis conclusions.

Finally, the paper develops personal asset pricing to show why benchmarking is still relevant even when the overlapping equilibrium model is not readily testable. By adopting personal asset pricing, investors can circumvent equilibrium non-testability by assess anomalies relative to investor specific evaluation criteria. The personal asset pricing approach gives an investor specific interpretation to standard anomaly results and provides unambiguous investment recommendations. When taken to its logical conclusion, investor heterogeneity, which is the fundamental building block of personal asset pricing, implies that active investing is not necessarily a zero-sum game for counterparties as is commonly assumed (French, 2008).

The premium investment framework and personal asset pricing encapsulate two related contribution topics. The premium investment framework argues that passive and active bets are complements. It then tracks the implications of the claim and shows how adopting the alternative view has led to speculative conclusions in the literature. Personal asset pricing contributes by giving a new interpretation of anomaly results under heterogenous preferences as a method to circumvent equilibrium non-testability. Moreover, it motivates the use of benchmarks in empirical settings and follows the implications of investor heterogeneity to the logical conclusion that active trades can be welfare improving for all counterparties.

2. Understanding anomalies

2.1. Anomalous in relation to what?

Over the past four decades, researchers have uncovered a variety of profitable rule-based active investment strategies. These strategies have been given a multitude of names depending on the setting in which they appear; anomalies, return predicting signals, factors, active bets and smart beta are just a few well suited and widely used names that this paper will also interchangeably adopt. And when it comes to academic work, these anomalies are truly widespread (Green, Hand & Zhang, 2013; Harvey, Liu & Zhu, 2015).

The historical development of anomalies is inexorably intertwined with the empirical testing of the conditional CAPM. The returns to portfolios created by sorting assets on some specific characteristic were considered anomalous because they were not associated with a commensurate rise in risk. In the CAMP, the beta of a security with respect to aggregate wealth is synonymous for risk. Consequently, the absence of beta was considered anomalous (Cochrane, 2011). In this respect, the first generation of anomalies was initially developed as CAPM (with equity proxy) anomalies. Whenever researchers 'test' the CAPM with a proxy, they are testing if adding anomalies to the equity premium improves performance. Intuitively, time series alphas imply that the inclusion of anomalies will increase the risk-adjusted performance of a portfolio consisting of the right-hand side assets in a regression (Ferson, & Lin, 2014). Therefore, the historical development of anomalies suggests that they should be complements rather than substitutes to passive investments.

2.2. CAPM failures and the rising acceptance of data reduction techniques

Early tests of the CAPM discovered a wide variety of characteristic sorted portfolios that provide significant equity beta adjusted returns (Ball, 1978; Stattman, 1980; Banz, 1981; Basu, 1983; Rosenberg, Reid, & Lanstein, 1985; Chan, Hamao, & Lakonishok, 1991). This collection of first generation anomalies was eventually subsumed by two characteristics: (1) book to market (value) and (2) market capitalization (size) (Fama & French, 1992). Not long after, momentum joined center stage (Jegadeesh & Titman, 1993; Grundy & Martin, 2001) to form the Fama-French (FF)-Carhart model (Carhart, 1997).

Regardless of Roll's critique (Roll, 1977) and the inherent non-testability of the CAPM[2], over the years, researchers have used a variety of approaches in the hope that they will resurrect the underlying logic and

[2] The fact the true market portfolio cannot be observed implies that tests are conducted on the mean variance efficiency of the proxy portfolio. Regardless of the inherent inability to observe the true market portfolio, a correlation higher than

intuition of the model. Their hope was to obtain a positive coefficient on market betas and insignificant coefficients (alphas) for all the other factors (anomalies) in cross-sectional (time series) tests. This strand of work has mainly centered on better Bayesian estimates of betta (Vasicek, 1973; Karolyi, 1992), better proxies for the portfolio of aggregate wealth (Stambaugh, 1982; Jagannathan & Wang, 1996) and beta conditioning (Ferson & Harvey, 1999; Avramov & Chordia, 2006; Lewellen & Nagel, 2006).

Even when successful at resurrecting the significance of beta, most studies failed to account for the anomalous alphas associated with core anomalies. For example, Avramov and Chordia (2006) find that the conditional CAPM cannot explain size, book to market and momentum. Similarly, Lewellen and Nagel (2006) show that if the conditional CAPM truly holds, then deviations from the unconditional CAPM should be smaller than the ones observed empirically. Since early tests of the conditional CAPM failed to convincingly account for the main factors (size, value and momentum), they became accepted in the literature and formed a new benchmark. Latter studies had to face a higher hurdle; they needed to remain significant after controlling for the FF+ Carhart factors.

The resulting empirically motivated FF+ Carhart model did not last long. An extensive second generation of anomalies soon followed (Sloan, 1996; Pontiff & Woodgate, 2008; Ang, Hodrick, Xing & Zhang, 2006; Asness, Frazzini, & Pedersen, 2015) and their excess returns could not be explained by a proportional rise in value, size, market or momentum betas. Consequently, they were considered anomalous also in relation to the benchmark anomalies. And the ink was not yet dry on the five-factor model (Fama & French, 2015), when a third generation of anomalies started to appear and gain in prominence. Even more disturbingly, the ability of the five-factor model to subsume or even dominate the existing multitude of anomalies remains highly questionable (Green, Hand, & Zhang 2014).

Empirically motivated models often do a reasonable job at data reduction, but they do not explain the underlying causes of anomalies. Resultantly, failure to resuscitate the CAPM has separated investigations into data reduction techniques that summarize signal information in a concise format useful for out-of-sample prediction but devoid of economic content (such as the Fama-French models or models based on principal component analysis) and investigations that try to explain anomalies. The next section summarizes the later strand of work which is of more interest in this paper.

2.3. Revisiting anomaly explanations

Our inability to subsume anomalies via CAPM modifications has driven the emergence of five strands of anomaly explanations. They are broadly group into (1) errors in estimation stories, (2), rational stories (3) behavioral stories, (4) implementation stories and (5) persistence stories.

0.7 of the chosen proxy with the true unobserved market would imply that a rejection of the proxy is a rejection of the true CAPM (Kandel & Stambaugh, 1987; Shanken, 1987).

Error in estimation stories focus on faults with the data or construction process. In effect, errors in estimation stories do not explain anomalies but question their existence. Consequently, the validity of error in estimation concerns can eliminate the predictive significance of anomalies and cast doubt on their true historical population relevance. To deal with error in estimation, out-of-sample tests in other markets (Fama & French, 2012), asset classes (Ang, Hodrick, Xing, & Zhang, 2009; Assness, Moskowitz, & Pedersen, 2013) or time frames (Davis, 1994; Chabot, Ghysels, & Jagannathan, 2014) are often employed. Alternatively, to account for extensive data-mining, it is also possible to adopt the complementary multiple hypothesis testing approach which requires profitable strategies to pass a higher t-statistic hurdle (Harvey, Liu, & Zhu, 2015). Assuming the evidence is strong enough to pass the existence hurdle, a key issue that needs to be examined is why the average investor prefers a particular side of a bet. Behavioral and rational stories are competing explanations for anomaly existence.

Rational stories rely on the key concept of high marginal utility states. Investors prefer a particular side of an active bet because it has positive realizations in periods of high marginal utility. In a sense, the CAPM is the ultimate rational story and it designates periods of low aggregate wealth as high marginal utility states.

Behavioral stories on the other hand, claim that the average investor chooses a particular side of the bet because: (1) he derives non-monetary utility from his bet (non-standard preferences), (2) is subject to some bias that leads him to wrongfully assess the probabilities of the bet (errors in expectations) (Daniel, Hirshleifer, & Subrahmanyam, 1998; Barberis, Shleifer, & Vishny, 1998) or (3) prefers a particular side of the bet due to constraints (rational but constrained; segmented markets) (Frazzini & Pedersen, 2013; Blitz & Vilet, 2007). Behavioral stories suggest that mispricing can be seen as a source of excess income for unconstrained investors that do not share the same non-standard preferences.

When it comes to non-standard preferences, the distinction between rational and behavioral becomes blurred. Is a preference for 'moral dividends' irrational? In a sense, all preferences are behavioral; people prefer safer cash flows over riskier cash flows. Having mentioned this caveat, it is useful to keep the accepted terminology 'rational' to refer to a commonly acceptable set of normative preferences. The distinction between rational and behavioral is particularly relevant when normative preferences are 'tested' against behavioral explanations in which investors make errors.

Assuming we have established that an anomaly exists, and we have a good story to justify its existence, an important question to consider is anomaly implementation. Even if anomalies exist and have a behavioral explanation, can they still be traded profitably? Implementation stories examine if executing an anomaly bet is impossible due to (1) transaction costs (Frazzini, Israel & Moskowitz, 2012; Novy-Marx & Velikov, 2016), (2) shorting and leverage constraints, (3) high variance and unfavorable higher moments[3] (Barroso & Santa-Clara,

[3] Refers to idiosyncratic crashes that do not occur in high marginal utility states. Consequently, crashes are not 'rational' explanation in the traditional sense. However, a valid implementation concern is whether idiosyncratic crash risk can be diversified properly in a portfolio.

2015) or (4) an inability to execute in real time (Lewellen, 2015). In this respect, implementation stories do not explain why anomalies arise which is what the rational-behavioral debate tackles. Rather implementation stories can motivate why anomalies persist; that is, market prices do no adjust as shrewd investors face difficulty placing offsetting bets.

Finally, persistence stories argue that anomalies are disappearing over time even though they may have existed historically. Such an outcome can be consistent with all the explanations for anomaly existence such as data-mining (error in estimation), market learning (behavioral), changeless in preferences (rational) and reductions in transaction costs (implementation). This paper focuses on the rational-behavioral debate and the next section lays out some general concerns about existing rational and behavioral theories.

2.4. Common difficulties to rational and behavioral theories

A commonly expressed concern with behavioral theories is that they have difficulty explaining the comovement *between* stocks in portfolios sorted on anomalous characteristics (Cochrane, 2011). Comovement and behavioral explanations are not inconsistent. Nevertheless, as Cochrane (2011) points out, theoretical behavioral models that motivate both alpha and the comovement pattern are rare.

Rational theoretical models on the other hand, have difficulty accounting for the lack of comovement *across* anomalies. The various active bets seem unrelated (or even negatively correlated) which makes fitting a single rational model difficult. To understand the intuition of this statement, consider the extreme case whereby two tradable portfolios with positive excess returns are perfectly negatively correlated. Such a setting would lead to an arbitrage opportunity which cannot be explained by any rational asset pricing model. More generally, the more independent anomalies researchers discover the closer we are moving from model mispricing to violations of the law of one price; that is, developing a rational model becomes more difficult as we move from a single anomaly with extensive idiosyncratic risk to uncorrelated multi-premium combinations.

This reasoning is well understood thought the Hansen-Jagganathan bound which places a high hurdle on rational theories (Appendix A develops the argument formally). If we find the magnitude of the equity premium puzzling (Mehra & Prescott, 1985), consider fitting something four times as large, which is what the passive-active combination in this paper implies. Yet, the literature is full of papers claiming they can rationally explain anomalies; usually by matching individual anomalies with specific environment mapping variables. The next section turns to pitfalls in reasoning and testing which can cause this overabundance of rational explanations.

2.5. Pitfalls in the empirical testing of rational models

Empirical investigations of rational models face two issues: (1) risk model mining and (2) testing of models in isolation. Risk model mining occurs when a researcher looks at anomaly performance to identify loss periods and then proceeds to test environment mapping variables that would define them as high marginal utility states. The anomalies aren't arbitrage opportunities. Inevitably there will be periods in which losses are made. What is really needed is a clear normative a-priori argumentation as to why a variable is used. We want to *test* explanations rather than forming them ex post for each premium by conveniently splitting the sample. Analogous to the standard arguments against data mining (Harvey, Liu & Zhu, 2015), testing for an array of rational anomaly explanations will inevitably lead to a significant finding by chance.

The problem of model mining is exacerbated by the second issue which is the tendency to test risk explanations in isolation. In the absence of market segmentation, pricing needs to hold for all valid assets ('for all assets i' in asset pricing models). We do not have one model to price stock A and another to price stock B. Consequently, rational explanations ought to explain all valid anomalies and portfolios of anomalies (as the HJ bound also implies). At the very least, new rational models should not make other anomalies even more anomalous; that is, they should not prescribe expected returns for other anomalies that are even further from observation.

A valid out-of-sample test for existing premium specific rational explanations is their ability to explain alternative anomalies. Following this reasoning, the paper show that recessions and crashes are anomaly specific which is contrary to previous conclusions in the literature (Chordia & Shivakumar, 2002; Barroso & Santa-Clara, 2015; Daniel & Moskowitz, 2016). The empirical investigation demonstrates the importance of out-of-sample testing for new rational explanations. Assuming testing is properly executed, the next section turns to general concerns about the inherent testability of rational and behavioral explanations.

2.6. Obstacles to equilibrium testing

Currently, the biggest stumbling block for anomalies is our inability to determine whether they are rational or behavioral in an equilibrium context. The intuition of the impasse can easily be understood though the following simple setting. Suppose that the true data generating process follows:

$$(1) \qquad R_{it} = \underline{\alpha}_i + \lambda_i R_{qt} + \epsilon_{it} \quad \text{for all } i$$

Where R_{it} and ϵ_{it} are respectively the the excess return and disturbance term for asset i in period t. R_{qt} is the true equilibrium price determinant and $\underline{\alpha}_i$ is asset specific mispricing.

A rational benchmark would imply that the true determinant (R_{qt}) captures a risk that investors should normatively price. Above and beyond that, and for obvious reasons, a rational benchmark implies the absence of asset specific mispricing ($\underline{\alpha}_i = 0$).

A behavioral benchmark can be modeled as: (1) systematic mispricing, by assuming that R_{qt} is behavioral (like sentiment)[4], or as (2) asset specific mispricing, if alphas in the true model are not zero ($\underline{\alpha}_i \neq 0$). In fact, there is nothing excluding the possibility that mispricing is both systematic and asset specific.

[4] See for example Baker, & Wurgler (2006), Stambaugh, Yu, & Yuan (2012) or Jacobs (2015) for empirical evidence on sentiment and anomalies.

Suppose in practice researchers run a regression of the form:

$$(2) \qquad R_{it} = \alpha_i + b_i R_{zt} + \acute{\varepsilon}_{it}$$

where they utilize a different benchmark and discover positive and significant alphas.

Claiming equilibrium mispricing in this setting is problematic given that there is always the possibility that researchers are omitting the true rational factor which if included would annul anomaly alphas. Stated alternatively, alpha and the benchmark are simultaneously tested (Fama, 2014). Unfortunately, the joint hypothesis problem is not the only obstacle to resolving the rational-behavioral debate.

To make things worse, absence of alphas does not imply rational pricing as the utilized benchmark could also capture systematic mispricing[5]. Therefore, even if we discover a benchmark that annuls anomaly alphas we would still be unable to settle the rational-behavioral debate. The fact that a common factor can capture both rational and irrational pricing is the benchmark ambiguity problem.

To summarize, significant alphas do not imply mispricing (joint hypothesis problem) and absence of alphas does not imply rational pricing (benchmark ambiguity problem). For alphas to be informative about the rational-behavioral debate, we would have to know both the true benchmark *and* its nature! Understanding the nature of the true benchmark requires inferring the preferences and thought process of the representative investors. Are 'his' preferences 'rational'? Is there an error in his pricing process? As we presently do not know neither the true benchmark nor its nature, equilibrium interpretations of anomaly results are speculative.

Benchmark ambiguity is particularly problematic when explanations are reverse engineered, empirically motivated or lack economic content. Nevertheless, even explanations coming from 'deep theory' could have an ambiguous interpretation. For example, the pricing of recession risk can capture either a systematic error where agents wrongly expect bad times to last too long (behavioral story) or agents experiencing recessions as low consumption states (rational story). Given the testability issues of the rational-behavioral debate, the next section turns to personal asset pricing as a method to give alternative meaning to benchmarks.

2.7. Personal asset pricing

The controversy surrounding benchmarks has pushed researchers into widespread use of excess returns as an assessment metric. In addition, our inability to presently test the rational-behavioral debate does not change the fact that investors need to make choices regarding investments in anomaly replicating funds on a regular basis. And if alphas and other assessment criteria are not informative about equilibrium pricing, then what can they tell us?

[5] Kozak, Nagel & Santos (2017) present a similar argument in parallel work that strong comovement can be consistent with sentiment. In their case, the possibility is used as criticism to data reduction techniques. Here it is also argued that the issue is general to any benchmark.

To circumvent the issues inherent to the rational-behavioral debate, results can also be interpreted through personal asset pricing whereby the benchmark and evaluation criteria are personal, and alpha follows. As investors are aware of their own benchmark and performance assessment criteria, they can evaluate investment strategies relative to their own personal asset pricing model. Under this interpretation, the second specification (equation 2) is not an attempt to model the true data generating process but rather an independent predetermined personal benchmark. By fixing the benchmark, the approach circumvents the joint hypothesis problem and allows for unambiguous investor specific recommendations. In fact, even if we do end up knowing the true benchmarks and their nature, investor specific recommendations remain valid in the presence of investor heterogeneity.

Let us express personal asset pricing formally. Suppose the premium investor is a price taker and has a stochastic discount factor (SDF) independent of the data generating process; that is, assume his SDF is independent from the SDF of the representative investor and has the following beta representation:

$$(3) \qquad M_{t+1}^{p} = a - B_i R_{z,t+1}$$

Where $R_{z,t+1}$ is the relevant investor specific pricing factor and M_{t+1}^{p} is the resulting personal SDF. For example, if the benchmark portfolio $R_{z,t+1}$ captures personal wealth, then marginal utility will be high when personal wealth is low (a personal CAPM if you will). It follows (see Campbell, 2000 or Cochrane, 2009 for a more detailed derivation):

$$(4) \qquad E_t^{p}\left[R_{i,t+1}^{ex}\right] = b_i E_t^{p}\left[R_{z,t+1}^{ex}\right] \text{ for all i}$$

Where $E_t^{p}[R_{z,t+1}^{ex}]$ is the price of risk, b_i is the quantity of risk and $E_t^{p}[R_{i,t+1}^{ex}]$ is the conditional expected return for asset i in the personal model; the p here stands for personal, as in belonging to a specific investor. The quantity of risk (b_i) can also be interpreted as the regression coefficient of an asset's return on the return on the factor.

Assume a worst-case scenario whereby the SDF of the representative investor is rational (no equilibrium alphas) and different from the SDF of the premium investor and takes the form[6]:

$$(5) \qquad M_{t+1} = c - \lambda_i R_{q,t+1}$$

Analogous to before, we obtain:

[6] We can always construct a single (multiple factor) representation of the SDF when the law of one price holds (Hansen, & Jagannathan, 1991; Cochrane, 2009).

$$(6) \qquad E_t\left[R_{i,t+1}^{ex}\right] = \lambda_i E_t\left[R_{q,t+1}^{ex}\right] \text{ for all i}$$

The key ingredient is the difference between the SDF of the representative investor and the personal SDF regardless of their specific form. In practice, the preferences and benchmark of the representative investor are unobservable. This raises misspecification concerns when standard regressions are interpreted as attempts to model the data-generating process. However, prices and average returns can be observed. This allows investors to compare market expected returns to personal expected returns and to obtain investor specific alphas:

$$(7) \qquad E_t\left[R_{i,t+1}^{ex}\right] - E_t^p\left[R_{i,t}^{ex}\right] = \lambda_i E_t\left[R_{q,t}^{ex}\right] - b_i E_t^p\left[R_{Z,}^{ex}\right] = E_t^p\left[\alpha_i^p\right]$$

Where α_i^p is the alpha of investor p for asset i, and it captures the difference between the expected return for the representative investor and the personal investor. This is something we can take to the data. In this setting, $E_t^p\left[R_{Z,}^{ex}\right]$ is given while b_i can be estimated. Moreover, market expected returns ($E_t\left[R_{i,t}^{ex}\right]$) are observable assuming average realized returns are not systematically different from expected returns.

Defined in this manner, personal alphas provide unambiguous investment recommendations; that is, we can provide partial normative statements with the new alpha interpretation. Regressions defined using investors' benchmark do not suffer from misspecification concerns as the preferences of a single investor can easily be elicited. Finally, even if the SDF of the representative investor becomes observable, the conclusions continue to hold under fixed heterogeneous preferences.

The idea that investors are heterogeneous and care about different attributes of financial assets is intuitive. For example, investors could care about different covariances as they must hold fixed non-tradable assets. Differences in preferences can also manifest themselves as higher marginal utility from returns in specific states of the world. For example, investors could care more about returns in recessions due to their labor income risk. Even non-standard preferences, such as a preference for socially responsible firms and moral dividends, can cause differences between personal and equilibrium valuations. For example, some ethical investors would not hold 'sin stocks' regardless of market prices (see Renneboog, Horst, & Zhang 2008 for a review on SRI investing). As these examples illustrate, personal asset pricing and heterogeneous preferences are inherently neither rational nor behavioral but can be consistent with both.

The use of investor specific criteria is not novel. On the contrary, the approach is quite intuitive and often implicitly used in practice when developing client specific solutions. This is what practitioners are doing when regressing on clients' portfolios. A similar approach can also be found in Ferson & Lin (2014) who derive investor specific alphas in the SDF setting as a reliable buy/sell signal in the presence of investor disagreement. In their case, they use the intuition to analyze fund flows. In this paper, personal asset pricing is used to understand and interpret anomaly results as well as to motivate the use of benchmarks in the presence of non-testability.

Personal asset pricing is not without its own shortcomings. A key valid drawback of personal asset pricing is that it provides investor specific conclusions. Every preference structure requires a new investigation. The next section continues exploring some of the unique concerns specific to personal asset pricing.

2.8. When will anomalies be risky under personal asset pricing?

Having analyzed anomalies from a personal perspective, a pressing question remains; what are the conditions required to make investors worse-off on average when they make anomaly investments? This section derives the consequences of completely ignoring the pricing process of the representative investor and the conditions under which they become relevant.

When will alphas disappear? If asset specific mispricing in the data-generating process is zero ($\underline{\alpha}_i = 0$)[7], then personal alphas are zero when the SDFs are equivalent:

$$\text{(8)} \qquad \text{If } M_t^p = M_t \text{ then } E_t^p\left[\alpha_i^p\right] = 0$$

This can occur as a structural break in preferences ex-post because of the interdependence of utility functions between the representative investor and the personal investor. Analyzed from this angle, alpha decay can be interpreted as an alignment between the preferences of the personal and representative investor.

For the active trade to be utility reducing for the buyer, we need the sign of alpha to switch in our beta representation (assuming λ_i is positive). Stated alternatively, we would also need the personal investor to demand a bigger premium for bearing factor risk than the representative investor. If the personal and representative investor agree about the quantity of risk, it follows:

$$E_t\left[R_{i,t}^{ex}\right] - E_t^p\left[R_{i,t}^{ex}\right] = \lambda_i E_t\left[R_{q,t}^{ex}\right] - \lambda_i E_t^p\left[R_{q,t}^{ex}\right] = E_t^p\left[\alpha_i^p\right]$$

$$\text{(9)} \qquad \text{If } E_t\left[R_{q,t}^{ex}\right] < E_t^p\left[R_{q,t}^{ex}\right] \text{ then } E_t^p\left[\alpha_i^p\right] < 0$$

If the personal investor demands a bigger premium for a risk, then he will find the asset overpriced for him. Alternatively, if the new personal SDF is a mix of the old personal SDF and the representative SDF:

$$\text{If } M_{t+1}^p = a - B_i R_{z,t+1} - \lambda_i R_{q,t+1}$$

[7] This is a weaker assumption than rational pricing which also requires absence of systematic mispricing ($R_{qt} \neq$ behavioural).

$$E_t\left[R_{i,t}^{ex}\right] - E_t^p\left[R_{i,t}^{ex}\right] = \left(\lambda_i E_t\left[R_{q,t}^{ex}\right] - \lambda_i E_t^p\left[R_{q,t}^{ex}\right]\right) - b_i E_t^p\left[R_{z,t+1}^{ex}\right]$$

$$E_t^p\left[\alpha_i^p\right] = \lambda_i\left(E_{t+T}\left[R_{q,t}^{ex}\right] - E_{t+T}^p\left[R_{q,t}^{ex}\right]\right) - b_i E_t^p\left[R_{z,t+1}^{ex}\right]$$

$$(10) \qquad E_t^p\left[\alpha_i^p\right] < 0 \text{ if } b_i E_t^p\left[R_{z,t+1}^{ex}\right] < \lambda_i\left(E_t\left[R_{q,t}^{ex}\right] - E_t^p\left[R_{q,t}^{ex}\right]\right)$$

Personal alpha will be negative when the difference in premiums demanded for the common risk cannot offset the unique personal risk of the asset.

To summarize, the true riskiness of anomalies from a personal perspective is in the probability that all three conditions are satisfied: (1) other investors must be pricing an additional unaccounted-for dimension of performance (regardless of whether it is rational or behavioral), (2) the premium investor must realize that he also cares about this other dimension of performance and that (3) he cares about the new dimension strongly enough such that the expected improvement on his initial dimension is not enough to compensate. If all conditions are satisfied, then the personal investor will experience an ex-post utility loss having realized he was trading on a pricing model inconsistent with his 'true' preferences. If we believe that the probability of a sign shifting structural break is negligible, then personal riskiness is non-existent. In short, completely ignoring equilibrium pricing and the representative investor is perfectly fine for agents who are confident about their preferences.

The final concern for personal asset pricing is measurement error. The personal investor has identical preferences as the representative investor. However, he estimated a positive alpha because he underestimated beta:

$$E_t^p[\alpha_i^p] = E_t\left[R_{q,t}^{ex}\right]\left[\lambda_i - \lambda_i^p\right]$$

$$(11) \quad 0 < E_t^p[\alpha_i^p] \text{ if } \lambda_i < \lambda_i^p$$

The price of risk in this example ($R_{q,t}^{ex}$) is common by the identical preference assumption. Measuring the true quantity of risk would have shown absence of alpha. Consequently, the measurement error problem can hurt the personal investor only to the extent that it induces him to excessively trade what he evaluated as incorrectly priced securities. Measurement error is more difficult to completely rule out (an equilibrium analog is testing whether beta conditioning can revive the CAPM). However, measurement error is also less problematic as it cannot flip the sign of gross alphas. We need a personal misspecification for the price and type of risk to obtain an alpha sign switch. Moreover, in frictionless markets with fixed preferences, if the investor places a positive probability that his estimated beta is the true beta, then personal alphas will be positive ex ante.

To summarize, ignoring equilibrium pricing can hurt the personal investor if: (1) he has a structural break in his preferences that puts him on the opposite side of the pricing process of the representative investor or (2) if it induces him to trade excessively properly priced assets due to an underestimation of the true quantity of risk. Having mentioned these unique risks, personal asset pricing provides a valid alternative for result interpretability given the testability issues inherent to the rational-behavioral debate. The next section explores the implications of heterogeneous agents on the equilibrium welfare pay-off to counterparties.

2.9. Heterogeneous agents and the zero-sum game property of anomalies

Anomalies represent zero-sum active bets across characteristics. These bets do not eliminate recession risk or directly generate wealth; rather, they shift it among market participants. Consequently, analogous to the way investors use active positions to earn an alpha, they can also use active positions to remove a risk. This zero-sum game property of active bets is not an empirical regularity but a fundamental property of markets. Theoretically, both outperformance and diversification of any risk can be achieved by a subset of investors at the expense of other market participants. However, anomalies are not necessarily zero-sum games when it comes to welfare as other market participants need not have a preference along the underperforming dimension. In fact, everyone may end up better off if agents have preferences across different dimensions. The following simple example illustrates.

Suppose there are two agents each holding two apples. If agent A trades away a fresh apple for a stale apple from agent B, and if both agents only care about apple freshness, then we have a zero-sum game in terms of both characteristics and welfare. Agent A got rid of the stale apple at the expense of agent B. This is the standard zero-sum game in terms of characteristics. With homogeneous preferences, it also implies a zero-sum game in terms of welfare among the counterparties[8].

However, if agent B *only* cares about another apple attribute such as color, then the exchange could be beneficial for both counterparties if the stale apple has agent B's preferred color. Both agents will find the apples personally mispriced relative to their preferences. Moreover, if the signals predicting freshness or color are changing over time, we will have continuous active trading that raises welfare. As the example demonstrates, showing this is trivial; for the two-agent case, simply consider a standard price-quantity diagram whereby supply and demand overlap for a range of prices. The goal of this stylized example is to show that anomalies need not be zero-sum games in terms of welfare for the counterparties as it is commonly believed in both academia and practice. The conclusion follows logically form the assumption of heterogeneous preferences.

[8] Focus is on the effect on counterparties. Active traders can also generate a positive social externality by contributing to price discovery (French, 2008).

In the chapters that follow, the paper empirically shows how the various discussed issues can impact analysis conclusions and our understanding of anomalies.

3. Methodological issues and methods to assess performance

3.1. Benchmark choice and bond-anomaly independence

The performance of anomalies will be assessed relative to a passive benchmark consisting of the equity, term structure and default premium. Under the personal asset pricing interpretation, this is an appropriate benchmark for investors who hold a passive portfolio of bond and stocks. One can also intuitively view benchmark betas as mapping high marginal utility states as periods when invested passive assets do poorly; that is, investors value assets more if they provide good returns when their personal wealth coming from passive assets is low. The choice to use the term structure and default premium is grounded in the work of Fama and French (1993) who find that they capture along with the equity premium most of the variation in portfolio returns of government and corporate bonds[9].

There are numerous alternative equilibrium ways to think about the relevance of contemporaneous anomaly independence from bond premiums. A simple way to understand the relevance is to revisit the intuition of the CAPM. Securities are risky if they have a high beta with respect to the portfolio of aggregate wealth. Since this portfolio is unobservable, it is common practice to use an equity index proxy. This proxy can also be expanded to include bond premiums (a quantitative approach to redefining the CAPM proxy, to include human capital in their case, can be found in Jagannathan & Wang, 1996). Consequently, if any of the anomalies correlate strongly with bond premiums, their excess returns can be considered rational in the sense that anomalies would be substitutes for bonds. A more general way to reason is to think of the term structure and default premiums as proxies for macroeconomic risk (for example, defaults tend to be clustered in periods of distress); i.e. as the true underlying determinants of systematic risk (Chen, Roll, R & Ross, 1986). Consequently, above and beyond the personal asset pricing interpretation, bond premiums also have a strong prior which ameliorates benchmark ambiguity concerns.

3.2. Alpha

Positive and significant time series alphas imply that including anomalies on the margin will increase the Sharpe ratio of a portfolio that contains the right-hand side benchmark assets (assuming their excess return is positive) (Dybvig & Ross, 1985). Therefore, for investors who already hold the passive premiums, adding an active strategy with a significant alpha will increase risk-adjusted performance. To assess alpha significance, the paper relies on the heteroskedasticity consistent standard errors of White (1980) (as in Stambaugh, Yu & Yan, 2012 and Jacobs, 2014). Conclusions are unaffected when using Newey West (1987) standard errors.

[9] Investors can get exposure to these premiums by buying (long term) corporate and government bonds.

3.3. The Sharpe and Sortino Ratio

Sharpe ratios are the most well-known metric for assessing portfolio performance and they take center stage. The statistical significance of Sharpe ratio differences is estimated using the Ledoit & Wolf (2008) bootstrap test[10] [11]. The paper also shows the annualized Modigliani & Modigliani (M2) performance measure which volatility matches strategies to a benchmark[12]. M2 has intuitive appeal, which is why it accompanies the results; however, it cannot qualify as a new measure as it is simply a restatement of the Sharpe ratio.

Anomalies are notorious for having highly asymmetric distributions (Figure 1)(see Eling & Schuhmacher (2007) for performance evaluation under asymmetric returns). To corroborate conclusions in the presence of non-normally distributed returns, the paper uses the Sortino ratio which is an alteration of the Sharpe ratio that uses downside deviation as the denominator. It is defined as:

$$S = \frac{R - T}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\min(0, R_t - T))^2}}$$

Where R is the average period return, T is the 'target' return, and R_t is the return in period t. With downside deviation, only returns falling below a certain threshold are considered risky. Since the paper investigates the performance of zero-cost portfolios, the natural target for downside deviation is zero. Sortino ratios also indicate how close investment strategies are to an arbitrage opportunity. Intuitively, a zero-cost portfolio that provides positive returns without downside deviation is the definition of arbitrage. Moreover, large Sortino ratios also indicate the absence of large losses which is especially relevant for compounded performance. These characteristics make Sortino ratios particularly well suited to the assessment of anomalies.

3.4. Recession Performance

Finally, the paper investigates anomalies' recession performance. The choice is intuitive as recessions lead to a drop in labor income which is a noteworthy risk for most investors. Moreover, consumption falls in

---

[10] An alternative approach to the use of Ledoit & Wolf's bootstrap procedure for differences in Sharpe ratios is the use of a GRS test (Gibbons, Ross, & Shanken, 1989) on all the estimated time series alphas to test for the improvement in *in-sample max Sharpe Ratios* of the augmented universe. The bootstrap procedure is preferable as it allows for greater flexibility in the type of portfolio improvements tested.

[11] Previous drafts arrived at equivalent conclusions using the Jobson & Korki's (1981) statistic augmented with Memmel's (2003) adjustment as well as Opdyke's tests (2007). Ledoit & Wolf's bootstrap test is preferred as it better accounts for non-normality and serial correlation in Sharpe ratios. Memmel (2003) does not account for either non-normal returns or serial correlation while Opdyke (2007) does not account for serial correlation. Serial correlation can be an issue for the denominator in the Sharpe ratio due to volatility clustering (Ledoit & Wolf, 2008).

[12] M2 is defined as: $\mathrm{M2} = \mathrm{R_p}\frac{\sigma_b}{\sigma_p}$ .Where $\mathrm{R_p}$ and σ_p are the annualized return and volatility of the portfolio and σ_b is the annualized volatility of the benchmark.

recessions (closer to habit; Campbell and Cochrane, 1995), which makes good returns particularly valuable (Cohrane, 2017). Consequently, recessions have a reasonably strong prior. It is therefore interesting to examine if they are a valid out-of-sample explanation or a premium specific risk.

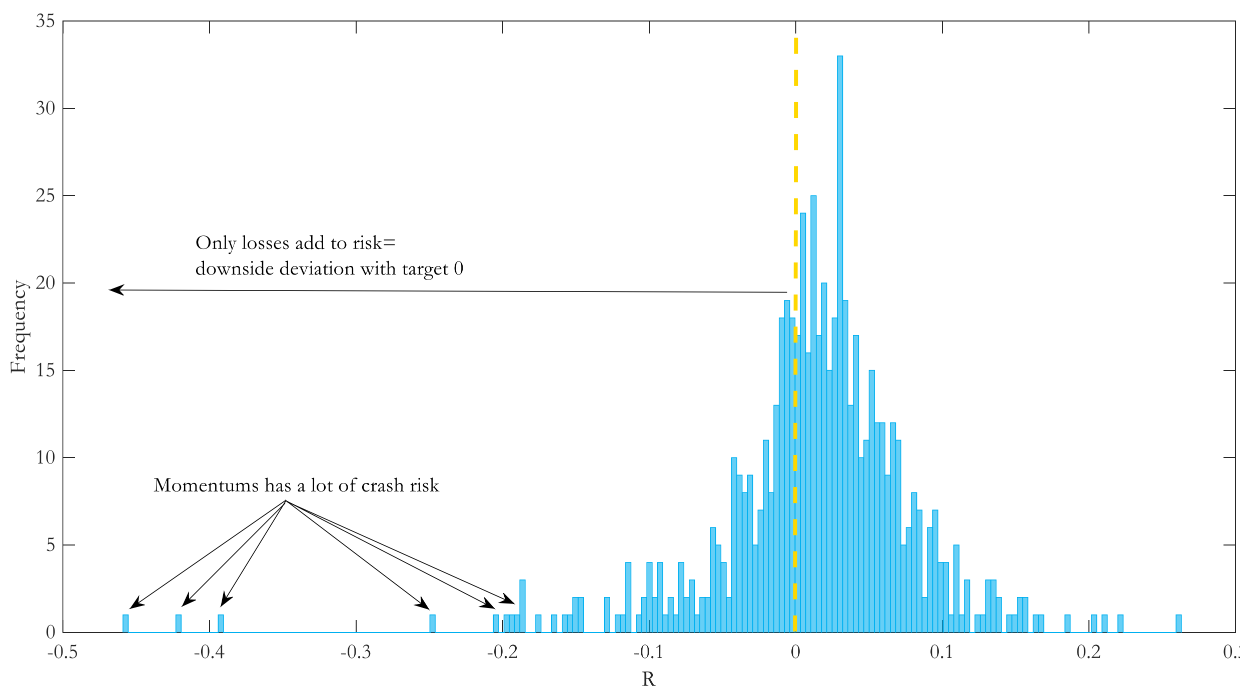

Figure 1: Momentum returns. Displays the frequency of monthly realized returns for momentum in the sample period spanning 07/1963-12/2014. Momentum has an asymmetric return distribution with negative skewness and high kurtosis.

3.5. Anomaly construction method

The most common methods for constructing and examining anomalies are (1) portfolio sorts and (2) Fama-MacBeth regressions. Fama-MacBeth regressions produce a time series of cross-sectional estimates (Fama & MacBeth, 1973) which have an interpretation of mean long-short hedge portfolios returns (Fama 1976, Chapter 9; Campbell 2014). The sorting approach on the other hand, places securities in portfolios based on the value of a specific characteristic. An anomaly is then constructed by going long the 'undervalued' portfolio and short the 'overvalued' portfolio.

Fama-MacBeth regressions and portfolio sorts have subtle differences, such as the effect of small stocks on the estimates (Fama & French 2008), nevertheless, they are conceptually equivalent given that portfolio sorts are the same as nonparametric cross-sectional regressions (see Cochrane (2011) for a visual illustration). It is common practice to apply both approaches when developing new anomalies and they generally provide equivalent conclusions (see the consistency of results across approaches in Fama & French, 2008). There is a very simple way to think about anomaly construction. All approaches make use of a panel data set containing an array of forecasting signal (Cochrane, 2011). These signals can be used to select securities and construct active portfolios.

The paper will use value weighted decile sorts, where applicable, as focus is placed on the performance of tradable portfolios. On average, equally weighted sorts provide stronger results (Green, Hand & Zhang, 2013) given that anomalies are often more pronounced in microcaps (Fama & French 2008). However, equally weighted (EW) portfolios are costlier to execute as they require rebalancing back to equal weights following monthly return realizations. Moreover, EW portfolios require a disproportionately high trading volume in small stocks. In fact, empirical studies show that EW portfolios have two to three times the transaction costs of VW portfolios (Novy-Marx & Velikov, 2015). Fama-MacBeth estimates have similar issues. They can be influenced extensively by microcaps, which are plentiful in the population and tend to take more extreme values in the characteristics (Fama & French 2008). Furthermore, the approach would also be inappropriate for anomalies such as betting against beta (Frazzini & Pedersen, 2014) which require further transformations before application.

The use of quintile rather than decile portfolios is another modeling alternative. Quintile portfolios have lower average idiosyncratic volatility due to the larger number of stocks per portfolio. However, quintile sorts have higher anomaly correlations due to the presence of a larger number of overlapping stocks. It is worth noting that overlapping stocks do not negatively influence the results above and beyond their impact on correlations. In fact, they can reduce the transaction costs of the overall portfolio when they give opposite trading recommendations.

3.6. Data Description

The main investigation uses US monthly return series for 13 zero-cost long-short decile portfolios constructed from 07/1963 until 12/2014. The paper gives all zero-cost long-short portfolios the general designation 'premiums', given that they all have positive average realized returns. The shorthand notation for the premiums used throughout the paper is as follows, MKT is the equity premium, GOV is the term structure premium (Asvanunt & Richardson, 2016), CORP is the default premium (Asvanunt & Richardson, 2016), SMB is the size (Banz, 1981; Fama & French, 1992, 2008, 2015), BTM is value (Rosenberg, Reid, & Lanstein, 1985; Chan, Hamao, & Lakonishok, 1991, Fama & French, 1992, 2008, 2015), RWM is profitability (Cohen, Gompers, & Vuolteenaho, 2002; Fama & French, 2008, 2015), CMA is investment (Fairfield, Whisenant, & Yohn 2003; Titman, Wei, & Xie, 2004; Fama & French, 2015), WML is momentum (Jegadeesh & Titman, 1993; Carhart, 1997; Fama & French, 2008), IVOL is idiosyncratic volatility (Ang, Hodrick, Xing & Zhang, 2006), QUAL is quality (Asness, Frazzini & Pedersen, 2015), BAB is betting against beta (Frazzini & Pedersen, 2014), AC is accruals (Sloan, 1996; Fama & French, 2008) and NI is net share issuance (Daniel & Titman, 2006; Pontiff & Woodgate, 2008; Fama & French, 2008). Return series were taken as given from previous work to enable ease of result replicability[13]. The construction procedure for each premium follows the specifics determined by the last paper cited. Interested readers can see the original papers for the details on anomaly construction. For ease of communication, the paper refers to MKT, GOV and CORP as the traditional premiums or passive bets. Remaining premiums are referred to as anomalies or active bets. All the anomalies, except betting against beta, are value weighted; betting against beta is constructed using the methodology of Frazzini & Pedersen (2013) whereby weights are assigned based on beta ranks. Recession data was obtained from the US national bureau of economic research (NBER). Recessions are defined as periods between peak and trough.

4. Results

4.1. Understanding the data: the premiums as stand-alone investments:

Table 1 summarizes premium information and reveals several interesting patterns. First and foremost, results reveal that premiums have highly asymmetric return distributions consistent with claims in the literature (Barroso & Santa-Clara, 2015; Daniel & Moskowitz, 2016). Skewness and kurtosis are on average quite large and the Jarque-Bera test rejects all the normality hypothesis.

However, skewness does not seem to have a universal sign deviation as five of the examined premiums have negative skewness while eight of them have positive skewness. Consequently, if skewness was a valid anomaly explanation, more than half of the active bets would become even more anomalous. Similarly, when it comes to kurtosis, it is difficult to explain why a premium goes in a particular direction as the opposite short-long bet can crash as well. Finally, and perhaps most importantly, higher moment explanations also assume that higher moment risks are not diversifiable. In the case of kurtosis, they assume that the multitude of premiums experience crashes simultaneously. This paper argues that the reduction of crash and skewness risk resulting from multi-premium combinations, coupled with the absence of a singular sign deviation of individual anomalies when it comes to skewness, are strong indications of the irrelevance of higher moment explanations for anomalies[14].

Second, except for the equity premium, all remaining portfolios have positive average returns in recessions. In fact, value, quality, idiosyncratic volatility, investments and net issuance, have higher than average returns during recessions (Table 1). This finding is in fact intuitive. In crisis, most stocks will lose in value. But highly volatile junk stocks can be expected to have an above average plunge. Since these stocks form the short positions in the respective anomaly portfolios, it is only natural that the accompanying anomalies do well in

---

[13] Special thanks to Fama, French, Frazzini, Pedersen, Asness, Asvanunt and Richardson for providing public access to their data.

[14] While misbehaving higher moments may indeed be a valid practical concern when it comes to anomaly bet implementation, they cannot "explain" why anomalies arise in the first place as they neither map high marginal utility states nor represent covariance with variables that do. If higher moment explanations were indeed a true risk explanation, then any long-short position with misbehaving higher moments, such as an industry bet, should have a premium. This reasoning inconsistent with both observation and theory.

recessions. However, recessions are also accompanied with an increase in volatility (see SD recessions in Table 1). Therefore, with investment being the exception, recession performance is not statistically significant.

Based on the initial data, it is difficult to argue that a particular premium is better than the equity premium as a stand-alone investment. Except for betting against beta, we cannot claim that anomalies offer a statistically significant improvement in Sharpe ratios (see Figure 9 appendix for visual illustration). In fact, most anomalies are close to the equity premium return/risk line. If investors formed a mental account of anomalies and evaluated them independently, then they may have arrived at an unfavorable view concerning anomaly performance. Rather than comparing the performance of the premiums as stand-alone investments, this paper turns to second portfolio construction stage and the reduction of premium specific risks[15]. This approach is more in line with the historical development of anomalies as complements to passive investments. With this end goal in mind, correlations and time series alphas are more relevant than premiums' standalone Sharpe Ratios.

Table 2 shows that average correlations among the various strategies are remarkably low which implies significant diversification potential. The choice of Pearson or Spearman correlations does not meaningfully alter the results. Among the strategies, the equity premium has the lowest average correlation. Bond premiums are also uncorrelated with each other and on average with the rest of the anomalies. In fact, relative to the equity premium, the correlation matrix suggests that anomalies show greater diversification potential than bonds.

Concerning recession performance, plotting the data illustrates a noteworthy preliminary pattern (Figure 2). The equity premium has contrasting performance relative to anomalies such as idiosyncratic volatility and quality specifically during recession periods.

[15] The first stage of portfolio construction is when we bundle individual securities into anomalies. The second stage is when we combine premiums into portfolios.

	MKT	GOV	CORP	SMB	BTM	RMW	CMA
R	0.51%	0.24%	0.13%	0.33%	0.51%	0.21%	0.46%
t-stat	*2.83*	*2.01*	*2.49*	*1.68*	*2.72*	*1.29*	*3.57*
R recession	-0.35%	0.45%	0.28%	0.15%	0.67%	0.19%	0.89%
t-stat	*-0.51*	*0.98*	*1.09*	*0.27*	*1.04*	*0.43*	*2.30*
SD	4.46%	2.99%	1.34%	4.83%	4.63%	3.97%	3.23%
SD recessions	6.43%	4.38%	2.44%	5.31%	6.11%	4.32%	3.69%
M2	6.26%	4.39%	5.39%	3.68%	6.02%	2.80%	7.87%
Sharpe Ratio	0.11	0.08	0.10	0.07	0.11	0.05	0.14
p-val Bootstrap		*0.58*	*0.78*	*0.37*	*0.94*	*0.43*	*0.67*
Skew	-0.54	0.34	-0.13	0.74	0.55	0.22	0.34
t-stat	*-5.5*	*3.5*	*-1.3*	*7.5*	*5.6*	*2.3*	*3.4*
Ex. Kurtosis	1.94	2.29	9.16	4.22	2.41	2.81	2.06
t-stat	*9.9*	*11.6*	*46.5*	*21.4*	*12.2*	*14.2*	*10.5*
JB test	127	147	2162	515	181	208	131
Max DR	56%	56%	21%	83%	52%	60%	32%
Target DD	3.08%	1.92%	0.90%	2.99%	2.82%	2.66%	1.96%
Sortino Ratio	0.16	0.13	0.15	0.11	0.18	0.08	0.24

Continued	WML	IVOL	QUAL	BAB	AC	NI	μ
R	1.32%	0.47%	0.42%	0.83%	0.40%	0.43%	0.48%
t-stat	*4.73*	*1.49*	*2.37*	*6.36*	*3.50*	*3.34*	2.95
R recession	0.40%	1.30%	0.88%	0.18%	0.11%	0.69%	0.45%
t-stat	*0.35*	*1.27*	*1.50*	*0.40*	*0.29*	*1.58*	0.84
SD	6.92%	7.85%	4.44%	3.24%	2.88%	3.18%	4.15%
SD recessions	10.9%	9.8%	5.6%	4.3%	3.5%	4.2%	5.45%
M2	10.9%	3.3%	5.2%	14.3%	7.7%	7.4%	6.56%
Sharpe Ratio	0.19	0.06	0.10	0.26	0.14	0.13	0.12
p-val Bootstrap	*0.34*	*0.53*	*0.83*	*0.03*	*0.69*	*0.79*	0.58
Skew	-1.51	-0.32	0.02	-0.61	0.55	0.22	-0.01
t-stat	*-15.3*	*-3.3*	*0.2*	*-6.2*	*5.6*	*2.2*	-0.10
Ex. Kurtosis	8.30	3.00	1.38	3.61	2.36	0.93	3.42
t-stat	*42.1*	*15.2*	*7.0*	*18.3*	*12.0*	*4.7*	17.4
JB test	2006	242	49	374	175	27	488
Max DR	80%	84%	56%	52%	26%	36%	53%
Target DD	4.94%	5.53%	2.91%	2.08%	1.71%	1.95%	2.73%
Sortino Ratio	0.27	0.09	0.15	0.40	0.24	0.22	0.18

Table 1: Data summary. MKT is the equity premium, GOV is the term structure premium, CORP is the default premium, SMB is the size, BTM is value, RWM is profitability, CMA is investment, WML is momentum, IVOL is idiosyncratic volatility, QUAL is quality, BAB is betting against beta, AC is accruals, NI is net share issuance and μ is the average value. Anomalies are constructed using monthly US data from 07/1963 until 12/2014. The target for the Sortino ratio and downside deviation is zero. M2 is the annualized Modigliani & Modigliani measure which volatility matches strategies to their benchmark (the equity premium). JB is the Jarque-Bera test for normality. The critical values for the test are 4.38 (10%), 5.88 (5%), 10.53 (1%). Maximum drawdown (Max DR) is the maximum percentage drop from a peak. T-statistics for the mean are computed using the heteroskedasticity consistent standard errors of White (1980). Statistical significance of differences in Sharpe ratios is calculated using the Ledoit-Wolf bootstrap test (with equity premium as benchmark).

Rho/Sig.	MKT	GOV	CORP	SMB	BTM	RMW	CMA	WML	IVOL	QUAL	BAB	AC	NI
MKT		0.00	0.00	0.00	0.39	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00
GOV	0.15		0.41	0.00	0.14	0.21	0.52	0.52	0.00	0.31	0.00	0.51	0.27
CORP	0.27	0.03		0.01	0.00	0.00	0.89	0.00	0.00	0.00	0.00	0.72	0.01
SMB	0.16	-0.13	0.11		0.00	0.00	0.00	0.09	0.00	0.00	0.00	0.00	0.00
BTM	-0.03	-0.06	0.15	0.34		0.00	0.00	0.00	0.89	0.00	0.00	0.14	0.45
RMW	-0.37	0.05	-0.17	-0.54	-0.38		0.00	0.00	0.00	0.00	0.17	0.40	0.00
CMA	-0.22	-0.03	0.01	0.12	0.46	-0.14		0.87	0.00	0.10	0.00	0.00	0.00
WML	-0.12	0.03	-0.14	-0.07	-0.16	0.15	-0.01		0.04	0.00	0.83	0.00	0.07
IVOL	-0.61	0.17	-0.15	-0.57	-0.01	0.58	0.20	0.08		0.00	0.00	0.00	0.00
QUAL	-0.46	0.04	-0.18	-0.61	-0.42	0.78	-0.07	0.15	0.66		0.70	0.01	0.00
BAB	-0.08	0.12	0.12	0.17	0.27	0.06	0.25	0.01	0.27	0.02		0.22	0.00
AC	-0.14	0.03	0.01	-0.15	0.06	0.03	0.17	0.17	0.18	0.10	0.05		0.01
NI	-0.38	0.04	-0.10	-0.45	-0.03	0.44	0.30	0.07	0.60	0.56	0.15	0.10	
Average	-0.15	0.04	0.00	-0.14	0.02	0.04	0.09	0.01	0.12	0.05	0.12	0.11	0.11

Table 2: Full sample unconditional Spearman correlation matrix. Reports the Spearman correlation coefficient calculated from monthly returns from 07/1963 until 12/2014. Numbers bellow the diagonal are the Spearman correlations. Numbers above the diagonal are p-values. Spearman correlation is less sensitivity to extreme values (Pearson provides similar estimates). MKT is the equity premium, GOV is the term structure premium, CORP is the default premium, SMB is the size, BTM is value, RWM is profitability, CMA is investment, WML is momentum, IVOL is idiosyncratic volatility, QUAL is quality, BAB is betting against beta, AC is accruals and NI is net share issuance.

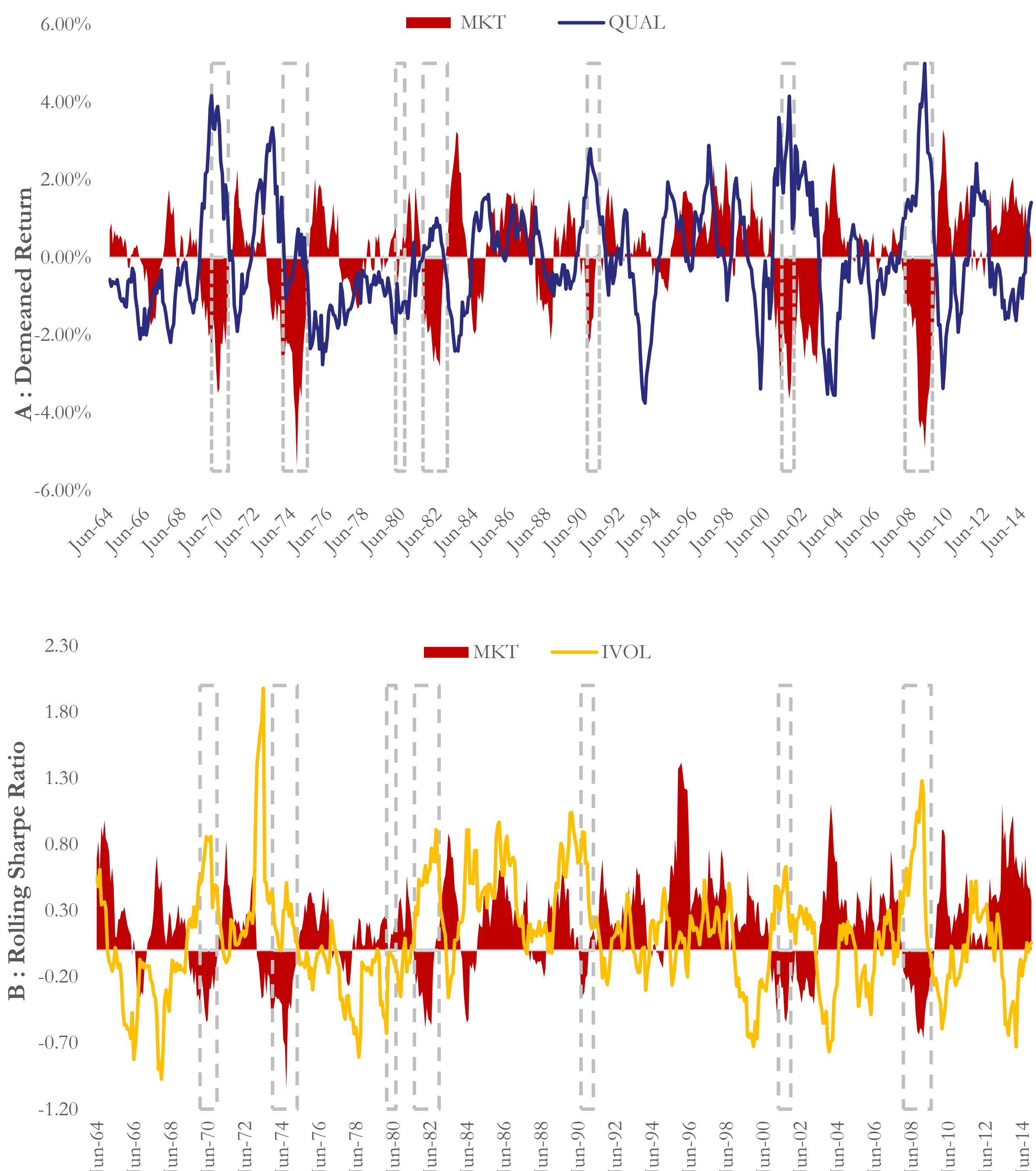

Figure 2: Panel A: Performance of Demeaned Characteristic Bets. Displays the average demeaned 12-month rolling performance of the equity premium (MKT) and quality (QUAL) whereby deviations below zero are periods when the crowded characteristic outperforms in the absence of a premium. **Panel B: Rolling Sharpe Ratios.** Displays the 12-month rolling Sharpe ratio between t and t-12 of the equity premium (MKT) and idiosyncratic volatility (IVOL). Gray rectangles indicate economic recessions as defined by NBER.

5. Time series regressions

In this section, the paper examines three related questions: (1) do anomaly alphas pass the new data mining adjusted t-statistic hurdle after controlling for the term structure and default loadings, (2) are term structure and default loadings economically relevant and statistically significant and (3) do anomalies have alphas in recessions controlling for their co-movement with bond premiums. All three related issues are addressed using two constant coefficient time series specifications. The results and specifications are shown in Table 3.

5.1. Alphas and loadings

The first specification (Table 3) contains the traditional premiums as benchmark assets. The loadings on these assets represent weights for a portfolio replicating anomaly performance. Exact replicability would imply the absence of alpha and the mean variance efficiency of a portfolio constructed with the passive benchmark assets (Gibbons, Ross & Shanken, 1989). Insignificant alphas would imply substitutability between the passive and active bets.

The first specification provides several noteworthy results. Looking at anomalies as a group, average loadings on all benchmark assets are very small. The average loading is -0.27 for the equity premium, 0.1 for the term structure premium and 0.03 for the default premium. Surprisingly, the average loading on the market is lower than the average loading on bond premiums. In other words, stock anomalies comove more with bond premiums than the equity premium.

As a group, anomalies are highly statistically significant as suggested by the high average alpha t-statistic of 3.86. This implies that loadings are not sufficiently large to annul anomaly significance. In fact, negative average loadings make average alphas (0.64%) larger than average anomaly excess returns (0.54%).

Looking at the individual regressions, we can see that between anomaly variation in loadings is considerable. The average loading for the equity premium is negative; in fact, except for size, all anomalies have negative equity loadings. For seven of the ten examined anomalies, equity loadings are statistically smaller than zero.

Anomalies based on characteristics associated with safety should comove more strongly with bonds. This hypothesis is supported by the data. As expected, IVOL has the strongest loading on the term structure premium. Similarly, quality, profitability and big (the reverse of size) have a statistically significant term structure loading. The result is in line with Baker and Wurgler (2012) who find that safe, large and profitable firms are more bond-like.

Default loadings are on average very small and are only economically and statistically significant for value and momentum (0.63 and -0.84 respectively). However, the sign of the loadings is opposite. Momentum looks more profitable while value looks less like an anomaly when default loadings are considered. For six of the anomalies examined, the term structure and default loadings go in the opposite direction.

Seven out of the ten anomalies examined, have an alpha t-statistic larger than 3. Value and profitability only pass the standard t-statistic hurdle of 2. And while some may argue that the data mining hurdles are too strict for early discoveries such as value (given that much of its return history is already out-of-sample), size is never reliably significant at any reasonable level.

Overall, the out-of-sample model tests shows that the equity, term structure and default premium are irrelevant for explaining anomaly returns (Appendix B shows that late anomalies are not linear functions of early anomalies). In fact, benchmarking makes things worse as alphas are larger than anomaly excess returns. Nevertheless, under personal asset pricing the goal is *not* to explain anomalies. The goal is to benchmark against invested assets to obtain unambiguous investment recommendations. Therefore, benchmarking on invested assets is necessary even when it adds to anomaly profitability. This is very different from the out-of-sample test of an equilibrium model interpretation under which the regressions show that the baseline premiums are unable to rationally explain (out-of-sample) anomalies.

5.2. Recession Alphas

The second specification expands the benchmark regressions with a dummy variable for recessions. The end goal of this modification is to examine whether anomaly alphas are negative or reliably smaller during recessions after controlling for the term structure and default loadings. Intuitively, positive alpha should be even more important during recessions if they are high marginal utility states. Controlling for bond loadings is relevant as we already know that bond premiums have good recession performance (Table 1). Consequently, it is prudent to rule out the possibility that recession alphas cannot be attributed to anomalies' bond-like features.

Alpha ($\dot{\alpha}$) can be interpreted as benchmark adjusted expansion returns, while $\dot{\alpha}$ - r represents benchmark adjusted recession returns (Table 3). The second specification reveals that anomaly alphas on average are not reliably lower in recessions (the average t-statistic on the recession dummy is -0.47). The exception is betting against beta. The recession alpha of BAB is lower than its expansion alpha and the result is marginally significant with a t-stat of -1.97. It seems that all of BAB's alpha is outside recessions. Stated alternatively, during recession, BAB returns are replicated well by the passive benchmark assets.

The estimated recession dummies, despite being generally insignificant, do suggest that anomaly alphas are meaningfully lower in recessions (approximately 43% lower than outside recession). However, there is no evidence that recession alphas are negative. In other words, anomalies still outperform in recessions but not as much.

5.3. Robustness check and the location of 'super anomalies'

Overall the results suggest that adding a (statistically significant) anomaly on the margin with improve the risk-adjusted performance of a portfolio consisting of the passive assets. The result also supports the view that passive investments (as represented by the equity, term structure and default premiums) and active investments (as represented by the anomalies) are not substitutes. If anything, as the following robustness check visually illustrates, they are complements.

As a robustness check, the equity, term structure and default premiums are amalgamated into an equally weighted benchmark portfolio; hereafter referred to as the traditional portfolio (TP). It should be first noted that the alteration gives equivalent conclusions concerning anomaly significance (the alpha t-statistics are shown on the right of Figure 3). Second, in addition to having positive alphas, anomalies have negative loadings relative to the traditional portfolio. For example, despite providing an average return of 0.94% per month, IVOL has a TP lamda (λ) of -1.59 and a multivariate MKT beta (ß) of -1.14. In fact, IVOL has more market hedging potential than a market short which has, by definition, a beta of -1 and an average negative monthly return of 0.51%. Resultantly, these 'super' anomalies can be used to simultaneously improve risk-adjusted performance and hedge. In fact, they act as superior substitutes to insurance as their negative covariance with passive assets prescribes negative excess returns. The result is 'super anomalous' as something that provides positive returns should not simultaneously hedge. An everyday example is being paid to take insurance on your house.

	α	ß	γ	η	ά	r
SMB	0.26%	0.19	**-0.24**	0.20	0.26%	0.01%
BTM	0.46%	-0.05	-0.08	**0.63**	0.46%	0.05%
RMW	0.35%	**-0.34**	**0.16**	-0.11	0.41%	**-0.38%**
CMA	0.55%	-0.20	-0.02	0.16	0.51%	**0.28%**
WML	1.49%	-0.19	0.16	**-0.84**	1.66%	**-1.17%**
IVOL	0.89%	**-1.14**	**0.60**	0.07	0.94%	-0.33%
QUAL	0.66%	**-0.49**	**0.17**	-0.23	0.66%	0.04%
BAB	0.81%	-0.09	0.12	0.24	0.95%	**-0.93%**
AC	0.43%	-0.10	0.04	0.09	0.50%	**-0.49%**
NI	0.54%	-0.30	**0.11**	0.09	0.55%	-0.03%
μ	**0.64%**	**-0.27**	**0.10**	0.03	**0.69%**	**-0.29%**

t-stat (HC se)	*t (a)*	*t (ß)*	*t (γ)*	*t (η)*	*t (ά)*	*t (r)*
SMB	1.37	3.62	**-4.17**	1.32	*1.27*	*0.02*
BTM	2.56	-0.77	-1.08	**3.35**	*2.40*	*0.08*
RMW	2.33	**-7.50**	**3.54**	-0.99	*2.55*	*-0.87*
CMA	**4.23**	-5.25	-0.41	1.42	***3.69***	*0.75*
WML	**5.73**	-2.01	1.56	**-2.07**	***6.38***	*-1.03*
IVOL	**3.71**	**-15.22**	**7.40**	0.29	***3.66***	*-0.42*
QUAL	**4.40**	**-10.69**	**3.00**	-1.56	***4.01***	*0.08*
BAB	**6.23**	-1.67	1.97	1.46	***7.07***	***-1.97***
AC	**3.58**	-3.12	1.08	0.91	***4.17***	*-1.31*
NI	**4.50**	8.72	**2.81**	0.94	***4.57***	*-0.07*
μ	**3.86**	**-3.39**	**1.57**	0.51	**3.98**	**-0.47**

Table 3: Bond-Anomaly independence and recession alphas. Uses monthly US data from 07/1963 until 12/2014. MKT is the equity premium, GOV is the term structure premium, CORP is the default premium, SMB is size, BTM is value, RWM is profitability, CMA is investment, WML is momentum, IVOL is idiosyncratic volatility, QUAL is quality, BAB is betting against beta, AC is accruals and NI is net share issuance. Recession is a dummy variable (1 for recession) obtained from NBER. Reported are the alphas and loadings from constant coefficient unconditional time-series regressions:

(1) $$R_{i,t} = \alpha_i + ß_i MKT_t + \gamma_i GOV_t + \eta_i CORP_t + \varepsilon_{i,t}.$$

(2) $$R_{i,t} = \acute{\alpha}_i + b_i MKT_t + \gamma_i GOV_t + \acute{\eta}_i CORP_t + r_i Recession_t + \acute{\varepsilon}_{i,t}.$$

where $R_{i,t}$ is the return of strategy i in month t. Loadings on MKT, GOV and CORP are similar across the two specifications and are therefore omitted for the sake of brevity. T-statistics are computed using the heteroskedasticity consistent standard errors of White (1980).

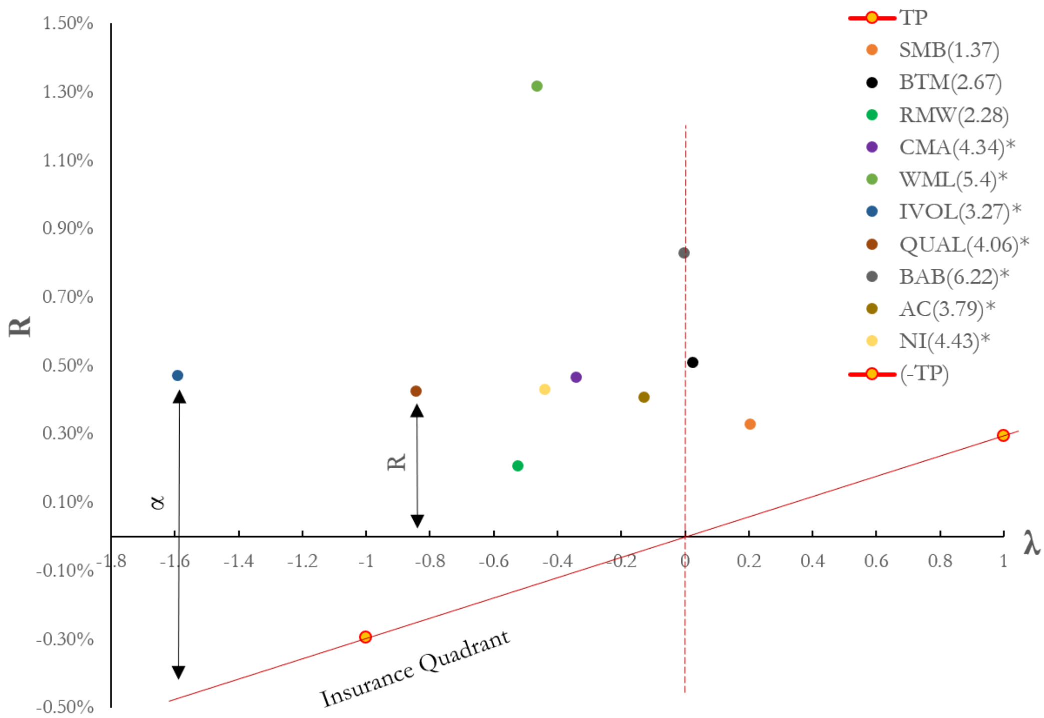

Figure 3: **Anomaly alphas.** Displays the performance of portfolios in return/loading space during the sample period spanning 07/1963-12/2014. TP is the traditional portfolio consisting of the equity, term structure and default premium. SMB is the size, BTM is value, RWM is profitability, CMA is investment, WML is momentum, IVOL is idiosyncratic volatility, QUAL is quality, BAB is betting against beta, AC is accruals and NI is net share issuance. Lambdas (λ) are calculated using constant coefficient unconditional time-series regressions:

$$R_{i,t} = \acute{\alpha}_i + \lambda_i\left(\frac{MKT_t + GOV_t + CORP_t}{3}\right) + \acute{\varepsilon}_{it}.$$

where $R_{i,t}$ is the return of a specific anomaly i in month t. Alpha t-statistics are reported in brackets. They are calculated using the heteroskedasticity consistent standard errors of White (1980). Asterisk accompany t-statistics higher than 3 (Harvey, Lie & Zhu, 2015).

6. Premium Portfolios

6.1. Portfolio choice

The investigation turns to multi-premium portfolios. The main objective is to investigate performance improvement across alternative performance dimensions such as crashes. The analysis makes use of five equally weighted portfolios. The traditional portfolio (TP) is a combination of the equity, term structure and default premiums. It proxies for investors' long-short version of a passive bond-stock investment. Size-value-momentum (SVM) represents a basic anomaly portfolio. The factor portfolio (FP) is an EW portfolio of all anomalies. It represents factor investing with an expanded anomaly universe. The mixed portfolio (MP) is a per premium EW portfolio. It represents a passive-active portfolio benchmark tilted towards active premiums. Finally, the balanced portfolio (BP) is a classification EW portfolio. It represents a balanced investment in active and passive investments.

An alternative to equal weights is the use of dynamic weights as prescribed by optimization techniques. However, past research indicates that alternative optimization methods underperform a naïve equally weighted benchmark out-of-sample due to estimation error (De Miguel, Garlappi & Uppal, 2009). Moreover, using equal weights is conservative as optimization techniques (especially in-sample) can further enhance the performance of an expanded universe and therefore even more strongly corroborate the conclusions.

6.2. The risk-adjusted performance of passive-active combinations

The multi-premium portfolio investigation corroborates the findings of the time-series regressions and suggested that passive-active combinations work well (Table 4). Both the premium portfolio (PP) and the balanced portfolio (BP) offer an economically sizable improvement in Sharpe ratios relative to the traditional portfolio (TP). M2 intuitively captures the magnitude of this improvement. The M2 of the balanced portfolio is 20.08% compared to only 7.83% for the traditional portfolio. The improvement in Sharpe ratios is statistically significant at the 1% level for the balanced portfolio and at 2% level for the premium portfolio. Combining passive and active bets gives a contrasting conclusion to comparing them as in Idzorek and Kowara (2013) for example.

Limits to arbitrage explanations tend to assume that large arbitrageurs endowed with superior information have principal agent problems with their investors. Short-term deviations that exacerbate mispricing harm arbitrageurs because they can cause fund withdrawal if they are misinterpreted as poor skill (Shleifer & Vishny, 1997). However, both standard deviations and downside deviations in the balanced portfolio are smaller than the traditional portfolio. Providers of capital will on average find these 'arbitrageurs' safer than passive bond-stock funds.

Transitioning from a factor portfolio to a balanced portfolio also results in an economically large and statistically significant improvement in Sharpe ratios. In addition, the significance of recession performance increases considerably in the expanded anomaly portfolio.

	Traditional	Factor Investing		Premium Investing	
	Passive	Active		Passive-Active	
	TP	SVM	FP	MP	BP
R	0.29%	0.72%	0.54%	0.48%	0.42%
t-stat	*3.58*	*5.65*	*6.38*	*8.05*	*9.12*
R recession	0.13%	0.41%	0.55%	0.45%	0.34%
t-stat	*0.37*	*1.03*	*1.74*	*2.10*	***2.01***
SD	2.05%	3.15%	2.09%	1.49%	1.13%
SD recessions	3.29%	3.76%	2.99%	2.04%	1.59%
M2	7.83%	12.64%	14.14%	17.79%	**20.08%**
Skew	-0.11	0.19	0.10	0.13	**0.04**
t-stat	*-1.07*	*1.95*	*1.04*	*1.36*	***0.38***
Excess Kurtosis	1.32	3.36	6.40	6.40	1.31
t-stat	*6.71*	*17.06*	*32.47*	*32.48*	*6.67*
JB test	44.8	288.4	1035.4	1036.8	43.3
Max DR	28.3%	25.0%	21.8%	14.9%	**8.3%**
Target DD	1.32%	1.84%	1.27%	0.85%	**0.60%**
Sortino Ratio	0.22	0.39	0.42	0.56	0.69
Sharpe Ratio	0.14	0.23	0.26	0.32	0.37
LW Bootstrap p-val (with TP)	0.225	0.173	0.149	**0.016**	**0.000**
LW Bootstrap p-val (with BP)	**0.000**	**0.000**	**0.014**	0.222	/
TP Beta	/	-0.08	-0.41	-0.08	0.30

Table 4: EW Portfolios. Uses monthly US data from 07/1963 until 12/2014. The traditional portfolio (TP) is an EW combination of the equity, term structure and default premiums. The size-value-momentum portfolio (SVM) is an EW portfolio of the size, value and momentum. The factor portfolio (FP) is an EW portfolio of the size, value, profitability, investments, momentum, idiosyncratic volatility, quality, betting against beta, accruals and net issuance. The mixed portfolio (MP) is a per premium equally weighted portfolio containing all active and passive premiums. The balanced portfolio (BP) assigns an EW weight to the traditional and factor portfolio (per classification EW). The market short balanced portfolio is the balanced portfolio that uses a market short instead of an anomaly short and low beta rather than (leveraged) betting against beta due to data limitations. The target for the Sortino ratio and downside deviation is zero. M2 is the annualized Modigliani & Modigliani measure which volatility matches strategies to a benchmark (the equity premium). JB is the Jarque-Bera test for normality. The critical values for the test are 10% (4.38), 5% (5.88), 1% (10.53). Maximum drawdown (Max DR) is the maximum percentage drop from a peak. T-statistics for the mean are computed using the heteroskedasticity consistent standard errors of White (1980). Statistical significance of differences in Sharpe ratios is calculated using the Ledoit-Wolf bootstrap test (LW Bootstrap p-val). The p-value for the Ledoit-Wolf Bootstrap TP and BP test is calculated with respect to the traditional portfolio (TP) and the balanced portfolio (BP) respectively (and the equity premium for the traditional portfolio).

6.3. Higher moments and the Sortino ratio

When it comes to skewness, we can see that multi-asset portfolios do not significantly deviate from normality. Excess kurtosis remains an issue especially for the factor portfolio. Nevertheless, in relative terms, the balanced portfolio has the smallest kurtosis. As figure 4 illustrates, and as the downside deviation statistic corroborates, the balanced portfolio does not have any large losses. The crash risk of individual premiums seems to be idiosyncratic.

The Sortino ratio provides an equivalent performance ordering as the Sharpe Ratio (Table 4). In fact, the Sortino ratio for passive-active portfolios is exceptionally high (0.69 for the balanced portfolio relative to 0.22 for the traditional portfolio). The improvement can be attributed to both an increase in returns and a decrease in downside deviations. In fact, as we move from the traditional to the balanced portfolio, downside deviations fall by more than a half.

The results show that the momentum crashes considered in Barroso and Santa-Clara (2015) and Daniel and Moskowitz (2016) are in fact premium specific. Complicated methods are not required to deal with crashes. Simple portfolio combinations work. More importantly, portfolios are a universal tool that can applied for any anomaly with crash risk concerns (not just momentum but also currency carry for example). The general implication is that rational explanations (or implementation concerns) should also be tested out-of-sample to examine if they are anomaly specific.

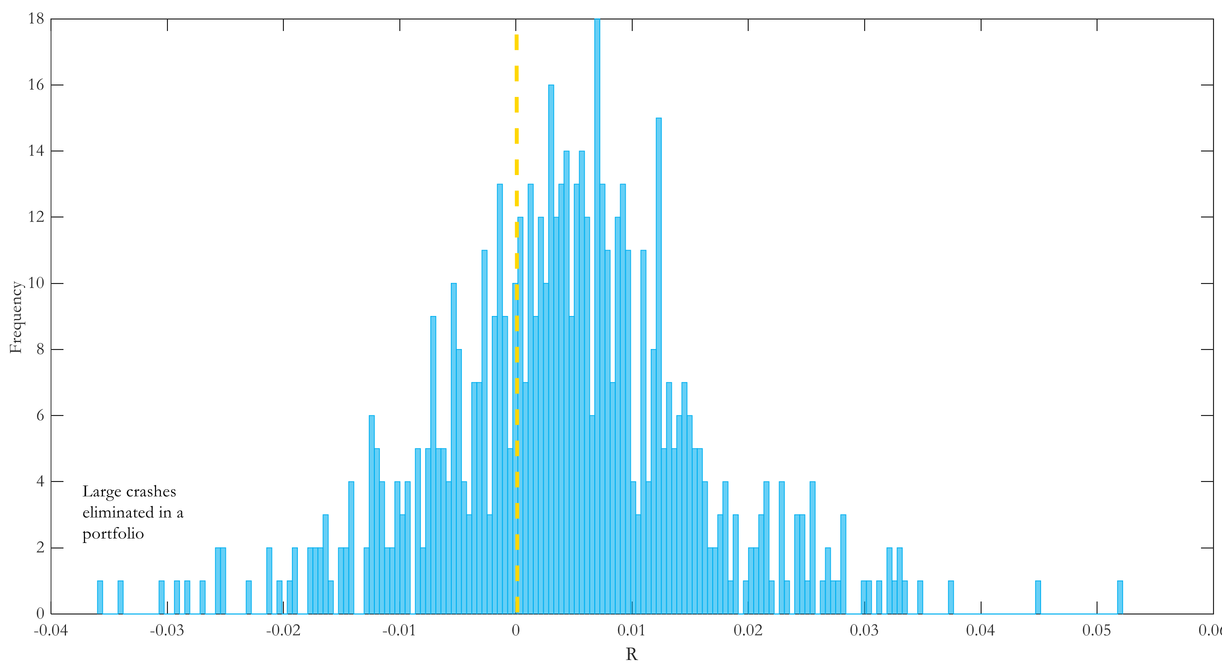

Figure 4: Balanced portfolio returns. Displays the frequency of monthly realized returns for the balanced portfolio in the sample period spanning 07/1963-12/2014. The balanced portfolio is constructed as an EW combination of the passive and active portfolio.

6.4. Recession performance

The high recession returns of the factor portfolio suggest that active premiums have a role to play in improving recession performance. However, the factor portfolio itself does not have significant recession returns. On the other hand, both the balanced and the premium portfolio have positive and statistically significant raw recession returns (t-statistics above 2). This suggests that passive-active combinations are specifically well suited to reducing recession volatility (and volatility in general) (Table 4). Overall, the result suggests that investors that hold passive-active portfolios can improve recession performance and even diversify recession risk.

The result again demonstrate that recessions are an idiosyncratic risk that cannot explain anomalies. Using the out-of-sample approach gives contrasting conclusions to Chordia and Shivakumar (2002) who investigate the recession performance of momentum in isolation.

6.5. Crashes, rolling betas, and performance persistence

The strong performance of the balanced portfolio can be attributed to the strong negative relation between active and passive strategies. The unconditional lambda of the factor portfolio, with respect to the traditional portfolio, stands at surprising -0.41 (see Figure 5 and the accompanying specification). There is a valid concern that correlations among asset classes and markets can increase over time (Bekaert, Hodrick, & Zhang, 2009) and rise sharply in recessions. Rolling least squares estimates suggest that this is not the case when it comes to active and passive bets. The beta of the factor portfolio on the traditional portfolio does not rise during crisis; in fact, it takes a favorable turn and sharply falls during the dot-com bubble before bouncing back close to unconditional levels. Despite the use of a limited set of observations for estimation, all the rolling beta estimates are reliably smaller than zero at a 99% confidence level; this implies a strong persistence to the negative relationship between active and passive bets. The finding also helps assuage beta measurement error concerns for personal asset pricing interpretations of anomaly results.

When it comes to crashes, Figure 10 in the appendix shows that there is no observation in the entire sample when both active and passive strategies crash simultaneously. This suggest that the reduction of crash risk achieved by passive-active combinations is not an artifact of the crash risk assessment metrics.

When it comes to outperformance persistence, Figure 6 illustrates two relevant points. Firstly, the outperformance of the balanced portfolio does not come from a sub-period as the balanced portfolio outperforms in most sample years and in every decade. Second, outperformance persistence also makes it particularly difficult to argue that anomalies are compensations for equilibrium risk. Looking at Figure 6 it is difficult to even hypothetically assign high marginal utility states that would make the balanced portfolio look worse-off. With the caveat that the magnitude of negative Sharpe ratios has an ambiguous interpretation[16],

[16] Do you want losses to be associated with more volatility or less volatility? More volatility reduces negative Sharpe ratios.

Figure 6 reveals that the single period in which the balanced portfolio meaningfully underperforms is in the years preceding the dot com bubble. However, arguing that anomalies exist because they compensate investors for bearing losses during bubble buildup periods is inherently difficult.

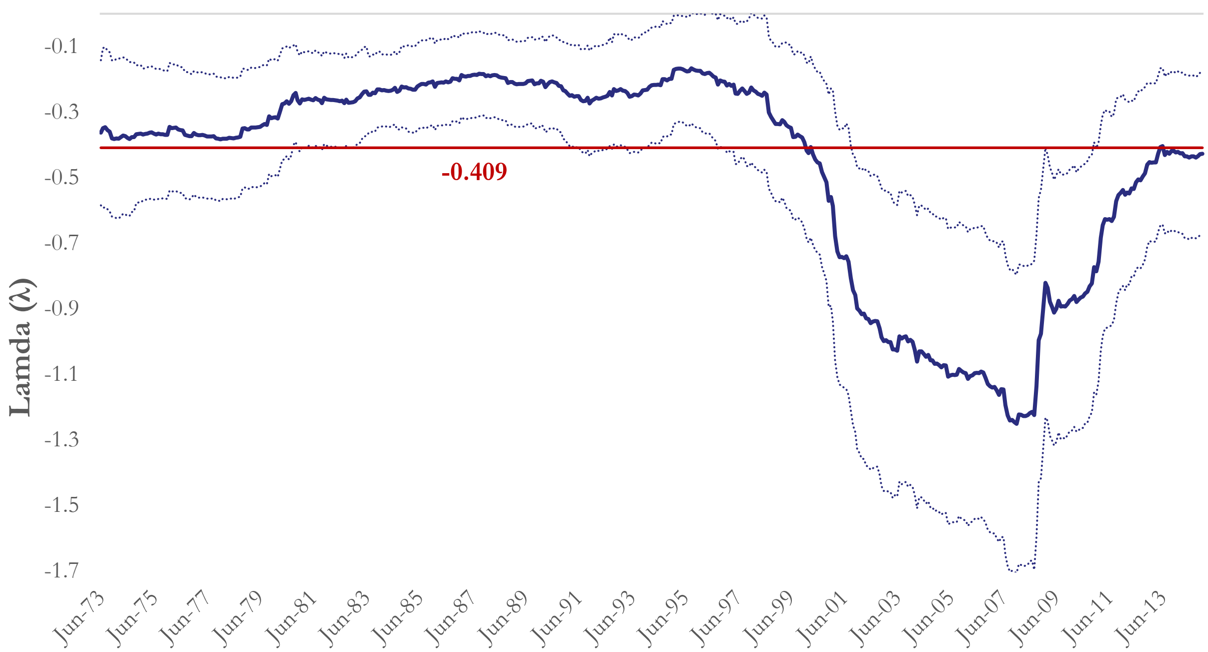

Figure 5: Rolling loading. Shows the 10-year rolling lamda of the factor portfolio on the traditional portfolio with the corresponding 99% confidence intervals. Rolling lamdas (λ) are calculated using:

$$R_{FP,t} = \acute{\alpha}_t + \lambda_t\left(\frac{MKT_t + GOV_t + CORP_t}{3}\right) + \acute{\varepsilon}_t$$

Time varying lamdas are shown in blue and are calculated with a data filter on a 10 year rolling window. Dotted lines represent rolling betas' 99% confidence interval. The full sample unconditional estimate is shown in orange. (TP) is constructed as an EW portfolio of equity (MKT), term structure (GOV) and default (CORP) premiums. The factor portfolio (FP) is constructed as an equally weighted portfolio of size, value, profitability, investment, momentum, idiosyncratic volatility, quality, betting against beta, accruals and net issuance.

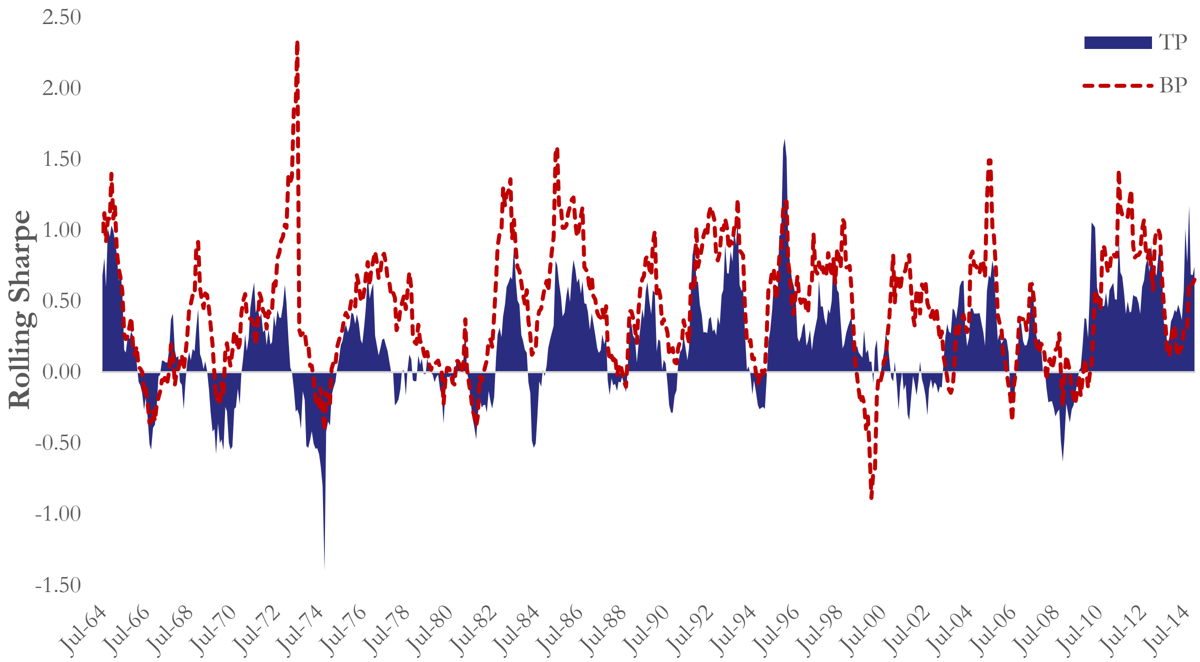

Figure 6: Outperformance persistence and the rolling annual Sharpe. Displays the 12-month rolling Sharpe ratio for the traditional portfolio (TP) and the balanced portfolio (BP). The traditional portfolio (TP) is constructed as an EW portfolio of equity (MKT), term structure (GOV) and default (CORP) premiums. The balanced portfolio is constructed as an EW combination of the traditional (TP) and factor portfolio (FP).

7. Robustness Check: Performance ordering with anomaly exclusion

7.1. Performance as a function of the number of anomalies

The previous section showed that anomalies, as a group, can improve upon passive portfolio performance. To investigate the sensitivity of the results to the specific choice of anomalies, Table 5 shows simulation results whereby all possible equally weighted sets of balanced portfolios are progressively constructed. The simulation has two goals: (1) to discover the cut-off point, in terms of number of anomalies utilized, where the worst performing balanced portfolio still outperforms the traditional portfolio and (2) to investigate the sensitivity of the results to the anomaly choice.

The results reveal that the average Sharpe ratio of the balanced portfolios is always economically larger than that of the traditional portfolio (Table 5). For example, a balanced portfolio constructed using four anomalies has an average Sharpe ratio that is twice as large as the Sharpe ratio of the traditional portfolio. As the number of anomalies increases, the improvement in performance becomes stronger (Figure 7). As expected, the use of more anomaly assets improves the distribution of achievable performance by removing premium specific risk.

The inclusion of any set of three anomalies to the traditional portfolio, *always* offers a better Sharpe ratio relative to the traditional portfolio. This mean that even if seven of the most performance enhancing anomalies were excluded from the analysis, the balanced portfolio would have still outperformed in terms of Sharpe ratios.

	Number of Included Anomalies					
	1	2	3	4	5	6
Max Sharpe	0.29	0.32	0.35	0.38	0.38	0.39
Average Sharpe	0.19	0.24	0.27	0.30	0.31	0.33
Min Sharpe	0.11	0.13	**0.15**	0.18	0.21	0.24
TP Sharpe	0.14	0.14	**0.14**	0.14	0.14	0.14
% (<Base)	30.0%	4.4%	**0.0%**	0.0%	0.0%	0.0%
Max Sortino	0.48	0.55	0.64	0.66	0.68	0.69
Average Sortino	0.30	0.40	0.46	0.51	0.55	0.59
Min Sortino	0.15	0.19	**0.23**	0.28	0.35	0.43
TP Sortino	0.22	0.22	**0.22**	0.22	0.22	0.22
% (<Base)	30.0%	4.4%	**0.0%**	0.0%	0.0%	0.0%
Max Recession Return	2.56	3.20	2.80	2.91	2.89	**2.84**
Average Recession Return	1.16	1.43	1.59	1.70	1.78	**1.85**
Min Recession Return	**0.39**	0.53	0.61	0.74	0.86	**1.08**
TP Recession Return	**0.37**	0.37	0.37	0.37	0.37	0.37
% (<Base)	**0%**	0%	0%	0%	0%	0%
% (<2)	20%	16%	24%	25%	28%	**30%**
# Portfolios	10	45	120	210	252	210

Table 5: Expanding anomaly universe. Displays the results for a simulation which calculates the Sharpe Ratio, Sortino Ratio and recession return t-statistics for all possible balanced portfolio combinations. The balanced portfolio contains an equal weight in the traditional portfolio (MKT+GOV+CORP), which is fixed, and an equal weight in the factor portfolio, which contains a varying number of anomalies. The number of included anomalies in the factor portfolio is increased at every step without replacement and every possible combination of the investigated measure is calculated. The table displays the maximum, minimum and average of the distribution of the investigated measure. % (<Base) shows the percent of balanced portfolios that fall under the traditional portfolio. % (<2) shows the percent of recession return t-statistics that are larger than 2. # Portfolios is the number of possible portfolio combinations without replacement $(\frac{n!}{k!(n-k)!})$ given a universe of anomaly assets.

Sortino ratios provide equivalent conclusions (Table 5). The maximum, average and minimum Sortino ratio, are monotonically increasing with the set of invested anomalies. Adding the three worst performing anomalies to the balanced portfolio provides a better Sortino Ratio relative to the traditional portfolio.

Given that time is fixed across portfolios, t-statistics represent scaled Sharpe ratios. Consequently, focus can be shifted from recession raw returns and Sharpe ratios to recession t-statistics. The goal is to additionally investigate the proportion of balanced portfolios that offer statistically significant recession performance.

The results reveal that recession return t-statistics display an even stronger patter than Sharpe and Sortino ratios (Table 5). The inclusion of any anomaly would improve the recession Sharpe ratio of the balanced portfolio above the level achieved by the traditional portfolio. Above and beyond that, recession return t-statistics larger than 2 are quite common. More than 30% of portfolios containing at least six anomalies exhibit robustness to recession performance (t-statistics larger than 2). The average t-statistics of recession performance is monotonically increasing from 1.16 for one anomaly, to 1.85 for six anomalies. The improvement can be considered sizable given that the recession return t-statistic of the traditional portfolio is only 0.37. Again, the improvement in recession performance is on top of the favorable recession effect of bond premiums. Overall, the results show that the conclusions are not dependent on the anomaly choice.

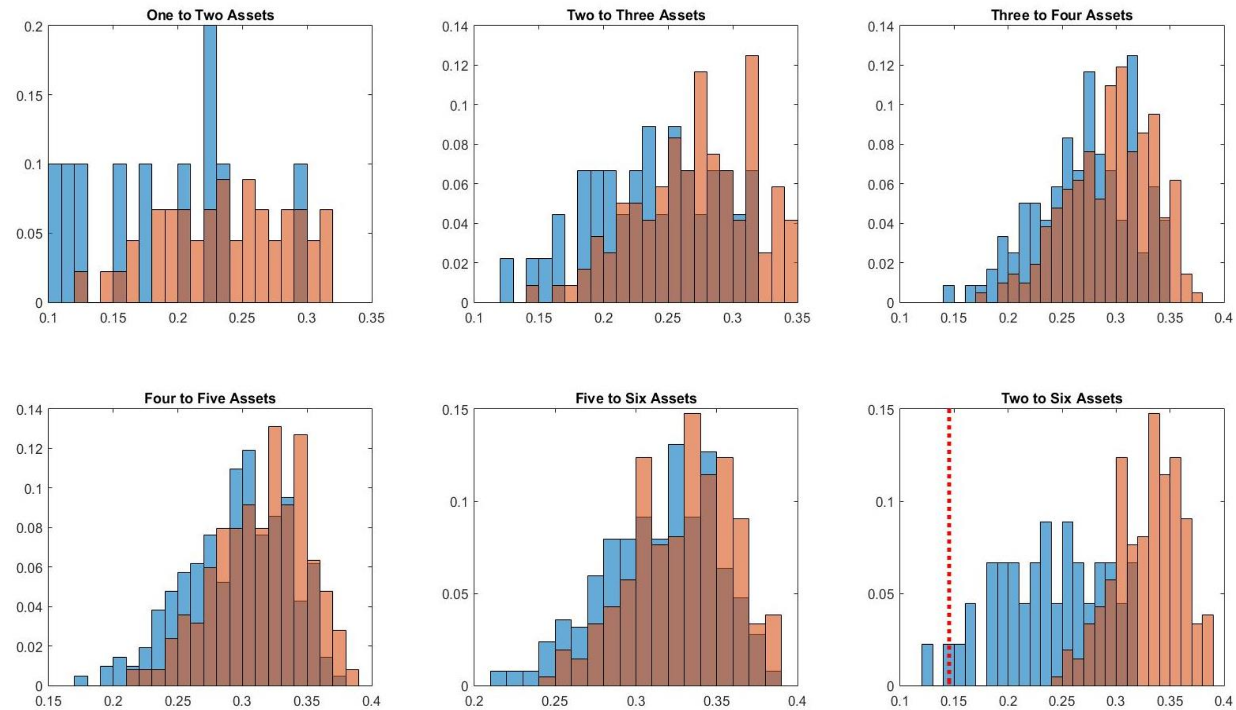

Figure 7: The distribution of Sharpe ratios as a function of the number of anomalies. Displays the distribution of balanced portfolios' Sharpe ratios as the number of anomalies included in the factor portfolio increases. The balanced portfolio is constructed as an EW of the traditional and factor portfolio. The Sharpe ratio of the traditional portfolio (0.14) is displayed with a red dotted line. The Sharpe ratios of portfolios constructed with a lower number of anomaly assets are displayed in blue. The Sharpe ratios of portfolios constructed with a higher number of anomaly assets are displayed in orange. The x-axis is the Sharpe ratio and the y-axis is the probability that a Sharpe ratio falls within a bin category.

7.2. Compounded performance

Passive-active combinations perform significantly better than passive portfolios across all three dimensions: (1) Sharpe Ratio, (2) Sortino Ratio and (3) recession performance. The results do not seem to suggest that this outperformance is limited to a sub-period or that it depends strongly on the choice of anomalies. As Figure 8 shows, the magnitude of risk-adjusted performance improvement is substantial. This improvement comes in addition to the fact that passive-active portfolios do specifically well in recessions. The figure also illustrates why active and passive bets should be joined (BP) rather than compared (TP vs FP).

Figure 8 also shows the performance of the counterparty portfolio and the high hurdle that rational explanations face. Equilibrium rational risk theories would have to explain the huge spread between the balanced portfolio (M2 = 20%) and the counterparty portfolio (M2 = -3.7%) whilst taking also into account the fact that the balanced portfolio does better in recessions and does not experience large crashes.

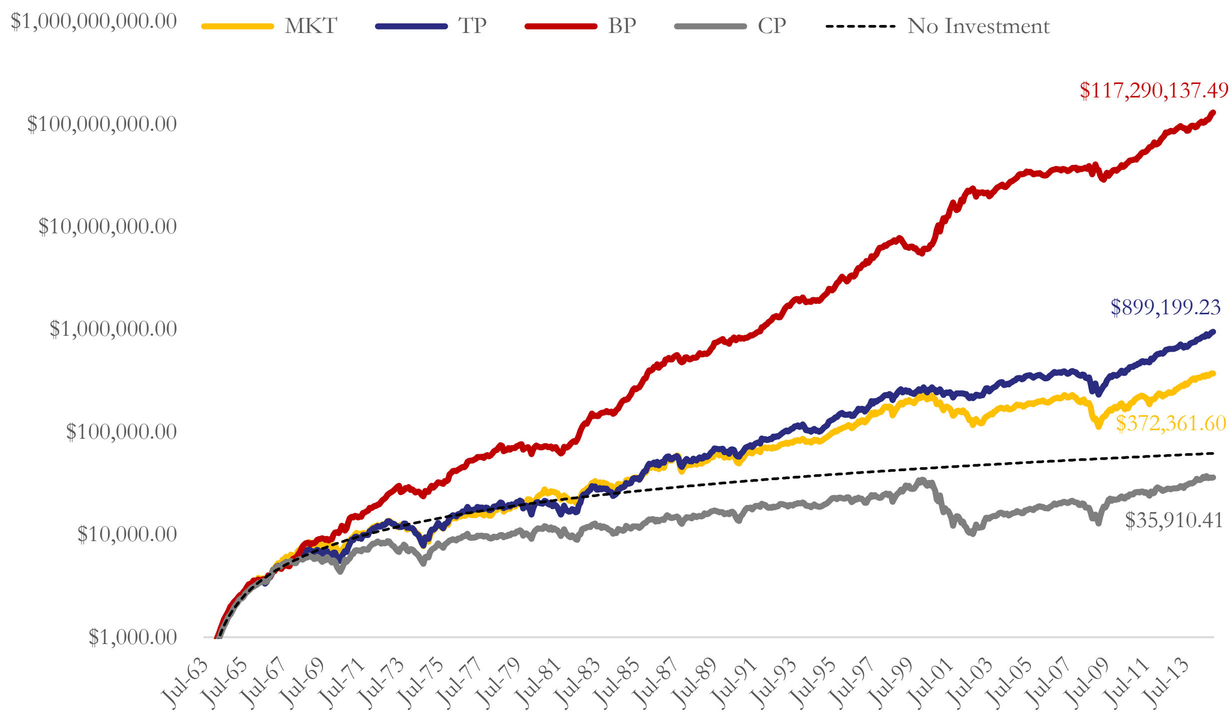

Figure 8: Compounded performance: dollar cost averaging the volatility matched portfolios. Displays end of sample compounded performance of volatility-matched dollar-cost-averaged portfolios that invest 100$ monthly and reinvests the proceeds; total sum invested over the full sample equals: number of periods times sum invested (618*100$=61800$). The dotted line represents the 'no investment' portfolio which in the case of zero-cost investing is the size of the position. Dollar cost averaging reduces the impact of early investing months. Portfolios are leveraged to have an equivalent full sample annual volatility as the equity premium (equivalent logic as M2). MKT is the equity premium. TP is the traditional portfolio which is an equally weighted combination of the equity (MKT), terms structure (GOV) and default premiums (CORP). The balanced portfolio (BP) is an EW combination of the traditional (TP) and factor portfolio (FP) (comprised of SMB, BTM, RMW, CMA, WML, IVOL, QUAL, BAB, ACC and NI). The counterparty portfolio (CP) takes the opposite bet by buying the traditional portfolio and selling the factor portfolio.

8. Alpha decay and the use of excess returns for anomaly assessment

8.1. Why is alpha decay relevant?

The paper argued that anomalies are a free lunch under a set of evaluation criteria and benchmark assets in frictionless markets under fixed preferences and a positive probability that the estimated beta is the true beta. If this is indeed the case, it is only prudent to examine anomaly persistence. Using equilibrium reasoning, and in the absence of frictions, the performance enhancing potential of anomalies should be eroded by the price readjustment pressure induced by investors who arrive at equivalent conclusions. If investors have a subset of the benchmark and evaluation criteria proposed in this paper, they will assess anomaly attractiveness in an equivalent manner and commit to premium investing. In the process, they will put price pressure on anomalies and re-price assets until passive-active combinations are no longer attractive. In the context of time series regressions, anomaly alphas should disappear.

8.2. Are excess returns or alphas the right performance measure?

The question of decay has received some attention recently. McLean and Pontiff (2016) find a degree of post publication *raw return* decay for an extensive array of anomalies. Focusing on raw returns (or Sharpe ratios) is not limited to investigations of anomaly persistence. Factor research also often focuses on the statistical significance of raw returns for assessing anomaly relevance. For example, Hou, Xue & Zhang (2014) declare 38 out of 80 anomalies to be insignificant based on the absence of raw returns. Similarly, Bali & Cakici (2008) criticize IVOL, among other issues, on its insufficient raw returns.

While the significance of raw returns may indeed be relevant for the assessment of anomalies as stand-alone investments, it is not the right approach from an investment perspective. The assumption underlying the use of raw returns is that the evaluated asset is the only one in the investment universe. Alternatively, in the presence of other securities, the procedure assumes that betas with respect to benchmark assets are zero. The argument is best understood looking back at Figure 3. The only setting in which alpha and excess returns are equivalent, is when an asset's beta to passive assets is zero.

Having this reasoning in mind, it is easy to see that anomalies can offer value even in the absence of significant raw returns if they have sufficiently high diversification potential as captured by their low betas to existing invested assets. Stated alternatively, distance to the zero-cost security market line is more relevant than distance from zero. In fact, if return decay is coupled with a strong enough decline in betas, then alphas could theoretically become larger. Alternatively, alpha decay can also come via an increase in benchmark betas (this can also be understood though Figure 3 as a shift in observations to the right). Therefore, raw return decay is only a proxy for the true measure of anomaly profitability decay.

The difference between alpha and raw returns is more than an inconsequential hypothetical possibility. Table 3 showed that betas with respect to benchmark assets can be negative. Moreover, as Figure 5 suggests, there has been a significant reduction of the factor to traditional beta in the past two decades (mainly concentrated during the dot com bubble). Consequently, researchers looking at excess return in this period, when betas declined sharply, could have derived erroneous conclusions regarding the profitability enhancing potential of anomalies.

To assess anomaly alpha decay, the paper implements a two-step approach. In the first step, five year rolling anomaly alphas are estimated using the bond-stock benchmark. In the second step, anomaly alphas are regressed on a time trend. Since alphas are the explained variables in the second regression, potential measurement error will not cause bias and inconsistency in the time trend estimates. Results are reported in Panel A of Table 6. The sign for the time trend is negative for six and positive for four of the examined anomalies. However, only two of the positive time trend coefficients are associated with a t-statistic above 2. More specifically, the alphas of quality and profitability have *increased* over time. Overall, the results do not support the existence of strong alpha decay across anomalies.

8.3. The rate of anomaly decay that would annul the validity of passive-active combinations

Finally, the paper estimates the rate of anomaly decay that would annul the usefulness of passive-active combinations. The procedure subtracts each month a constant from the raw returns of the factor portfolio. Following this modification, two tests are made: (1) an alpha significance test in a time series regression of the factor portfolio on the passive benchmarks (MKT, GOV, CORP) and (2) a bootstrap test on the difference between the decay adjusted balanced portfolio and the traditional portfolio. Results for both procedures are displayed in panel B of Table 6.

Subtracting a constant from raw returns is equivalent to alpha decay as it holds betas fixed. Intuitively, in the time series regressions, subtracting a constant from the left-hand side variable has no effect on the slopes but it reduces the intercept by an equivalent amount. Consequently, modeling raw return decay in this manner is equivalent to alpha decay.

The Sharpe ratio significance test reveals that the cutoff point above which anomalies no longer add value is close to a 3.5% percentage point decrease in anomaly returns and alphas. Such a decrease would yield a significant improvement in Sharpe ratios only at the 5% confidence level. The alpha significance test is even less strict. The inclusion of the factor portfolio spans the efficient frontier constructed by the benchmark assets by a statistically significant amount even following a 4.5% decay in alpha and returns.

Overall, the results suggest that a 50% drop in alphas is required to lose the statistical significance of the risk-adjusted improvement resulting from passive-active combinations. The results mean that either (1) wrongly assessed historical alphas by 50% or (2) future anomaly alpha decay of 50%, would make the passive-active strategy no longer statistically superior in terms of risk-adjusted performance relative to the passive benchmark.

Panel A			Panel B					
	Coef. Sign	t-stat (NW se)	Yearly Alpha Decay	Δ Sh	p-val	ά	t-stat (HC se)	t-stat (NW se)
SMB	-	-0.82	No decay	0.22	***0.00***	7.8%	***9.29***	***8.13***
BTM	-	-1.10	-1.0%	0.19	***0.00***	6.7%	***8.00***	***7.00***
RMW	+	**3.56**	-1.5%	0.17	***0.00***	6.2%	***7.43***	***6.50***
CMA	-	-0.13	-2.0%	0.15	***0.00***	5.8%	***6.86***	***6.00***
WML	-	-1.86	-2.5%	0.13	***0.00***	5.3%	***6.29***	***5.50***
IVOL	+	0.10	-3.0%	0.11	***0.01***	4.8%	***5.71***	***5.00***
QUAL	+	**2.02**	-3.5%	0.09	***0.04***	4.2%	***5.00***	***4.38***
BAB	-	-0.33	-4.0%	0.08	*0.09*	3.7%	***4.43***	***3.88***
AC	-	-0.89	-4.5%	0.06	*0.19*	3.2%	***3.86***	***3.38***
NI	+	1.78	-5.0%	0.04	*0.39*	2.8%	***3.29***	*2.88*

Table 6: Time trend estimates and alpha decay projections. Panel A shows the sign and significance of the time trend coefficient (λ) on rolling alphas.

$$(1) \qquad R_{i,t} = \hat{a}_{i,t} + \beta_{i,t}MKT_t + \gamma_{i,t}GOV_t + \eta_{i,t}CORP_t + \breve{\varepsilon}_{i,t}.$$
$$(2) \qquad \hat{a}_{i,t} = a_i + \lambda_i t + e_{i,t}$$

where i are the active premiums. T-statistics are calculated using Newey-West standard errors.

Panel B shows the rate of alpha decay that would annul the validity of passive active combinations. Holding beta constant, each month a fixed rate is subtracted from the return of the factor portfolio. Following the adjustment, two tests are executed: (1) the adjusted factor portfolio is EW with a traditional portfolio and the improvement and statistical significance of the combined portfolio relative to the traditional portfolio is calculated, and (2) a time series regression is executed of the adjusted factor portfolio on the passive benchmarks (MKT, GOV and CORP):

$$(3) \qquad R_{i,t} = \acute{\alpha}_i + \beta_i MKT_t + \gamma_i GOV_t + \eta_i CORP_t + \varepsilon_{i,t}.$$

where ά is the intercept in the regression. Standard errors for the regression estimates are calculated using either heteroskedasticity-consistent (t-stat (HC se)) or Newey-west standard errors (t-stat (NW se)). Δ Sh is the difference between the monthly Sharpe ratios of the balanced portfolio constructed with a decay adjusted factor portfolio and the traditional portfolio. The associated p-value is calculated using the Ledoit-Wolf bootstrap test.

9. Conclusion

The facts are simple. Despite having different term structure and default loadings, most anomalies continue to have high and statistically significant alphas which show no signs of decay. Sharpe ratio improvements relative to passive benchmarks are substantial and statistically significant. More importantly, adding anomalies also diversifies recession risk and helps control crash risk. The reduction of these risks suggests that they cannot account for anomalies and are not valid risk explanations. In short, not only are anomalies profitable, they also hedge. Robustness checks reveal that the conclusions are not sensitive to the choice of anomalies.

The findings show why rational anomaly explanations should be tested out-of-sample. Trying to rationally explain the performance of a single anomaly is likely to lead to idiosyncratic findings. The results also demonstrate how analysis conclusions can be influenced by adopting the erroneous substitutes interpretation to active and passive bets. Adopting this interpretation has led to widespread use of excess returns as a performance assessment measure. Similarly, it has resulted in comparison investigations between factors and traditional premiums. The empirical investigation shows how popular conclusions are altered by adopting methods more consistent with the complements interpretation.

In the presence of equilibrium non-testability, the interpretation of the results is motivated by personal asset pricing whereby investors consider their own benchmark and performance evaluation criteria when assessing anomalies. The approach helps circumvent the joint hypothesis problem by placing the individual investor at the center. Rather than assessing anomaly riskiness with respect to an unknown equilibrium asset pricing model, investors decide whether anomalies are personally risky. Under the personal asset pricing interpretation, investors who only care about performance across the dimensions examined, can think of the inclusion of anomalies to their passive portfolio as a free lunch in a frictionless world under fixed preferences and a positive probability that their estimated beta is the true beta.

REFERENCES

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The cross‐section of volatility and expected returns. *The Journal of Finance*, *61*(1), 259-299.

Ang, A., Hodrick, R.J., Xing, Y., & Zhang, X. (2009). High idiosyncratic volatility and low returns: International and further US evidence. *Journal of Financial Economics,* 91(1), 1-23.

Asness, C. S., Frazzini, A., & Pedersen, L. H. (2015). Quality minus junk. *Working paper*.

Asvanunt, A., & Richardson, S. (2016). The credit risk premium. *The Journal of Fixed Income* 26(3), 6-24.

Avramov, D., & Chordia, T. (2006). Asset pricing models and financial market anomalies. *Review of Financial Studies*, *19*(3), 1001-1040.

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross‐section of stock returns. *The Journal of Finance*, *61*(4), 1645-1680.

Baker, M., & Wurgler, J. (2012). Comovement and predictability relationships between bonds and the cross-section of stocks. *Review of Asset Pricing Studies*, *2*(1), 57-87.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, *9*(1), 3-18.

Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics,* 49(3), 307-343.

Barroso, P., & Santa-Clara, P. (2015). Momentum has its moments. *Journal of Financial Economics,* 116(1), 111-120.

Bouchaud, J. P., Krueger, P., Landier, A., & Thesmar, D. (2016). Sticky expectations and stock market anomalies. *Working paper.*

Campbell, J. Y. (2000). Asset pricing at the millennium. *The Journal of Finance*, *55*(4), 1515-1567.

Campbell, J. Y., & Cochrane, J. H. (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of political Economy*, *107*(2), 205-251.

Campbell, J.Y. (2014). Empirical Asset Pricing: Eugene Fama, Lars Peter Hansen, and Robert Shiller. *The Scandinavian Journal of Economics,* 116(3), 593-634.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, *52*(1), 57-82.

Carlin, B. I., Longstaff, F. A., & Matoba, K. (2014). Disagreement and asset prices. *Journal of Financial Economics*, *114*(2), 226-238.

Chabot, B., Ghysels, E., & Jagannathan, R. (2014). Momentum trading, return chasing, and predictable crashes *(Working paper No. w20660). National Bureau of Economic Research.*

Chan, L. K., Hamao, Y., & Lakonishok, J. (1991). Fundamentals and stock returns in Japan. *The Journal of Finance*, *46*(5), 1739-1764.

Chen, N. F., Roll, R., & Ross, S. A. (1986). Economic forces and the stock market. *Journal of Business*, 383-403.

Chordia, T., & Shivakumar, L. (2002). Momentum, business cycle, and time‐varying expected returns. *The Journal of Finance*, *57*(2), 985-1019.

Cochrane, J. H. (2009). *Asset Pricing:(Revised Edition)*. Princeton university press.

Cohen, R. B., Gompers, P. A., & Vuolteenaho, T. (2002). Who underreacts to cash-flow news? Evidence from trading between individuals and institutions. *Journal of Financial Economics*, *66*(2), 409-462.

Daniel, K., & Moskowitz, T. J. (2016). Momentum crashes. *Journal of Financial Economics,* 122(2), 221-246.

Daniel, K., & Titman, S. (2006). Market reactions to tangible and intangible information. *The Journal of Finance*, *61*(4), 1605-1643.

Daniel, K., Hirshleifer, D., and Subrahmanyam, A. (1998) Investor psychology and security market under- and overreaction, *Journal of Finance* 53, 1839–1885.

Davis, J. L. (1994). The cross‐section of realized stock returns: The pre‐COMPUSTAT evidence. *The Journal of Finance*, *49*(5), 1579-1593.

DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?. *Review of Financial Studies*, *22*(5), 1915-1953.

Dybvig, P. H., & Ross, S. A. (1985). Differential information and performance measurement using a security market line. *The Journal of Finance*, 40(2), 383-399.

Eling, M., & Schuhmacher, F. (2007). Does the choice of performance measure influence the evaluation of hedge funds?. *Journal of Banking & Finance*, *31*(9), 2632-2647.

Elton, E. J., Gruber, M. J., Brown, S. J., & Goetzmann, W. N. (2009). *Modern portfolio theory and investment analysis*. John Wiley & Sons.

Fairfield, P. M., Whisenant, J. S., & Yohn, T. L. (2003). Accrued earnings and growth: Implications for future profitability and market mispricing. *The accounting review*, *78*(1), 353-371.

Fama, E. F. (2014). Two pillars of asset pricing. *The American Economic Review*, *104*(6), 1467-1485.

Fama, E. F., & French, K. R. (1992). The cross‐section of expected stock returns. *the Journal of Finance*, *47*(2), 427-465.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3-56.

Fama, E. F., & French, K. R. (1998). Value versus growth: The international evidence. *The journal of finance*, *53*(6), 1975-1999.

Fama, E. F., & French, K. R. (2012). Size, value, and momentum in international stock returns. *Journal of Financial Economics*, *105*(3), 457-472.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1-22.

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *The Journal of Political Economy*, 81(3), 607-636.

Fama, Eugene F., (1976). *Foundations of Finance,* Basic Books.

Ferson, W. E., & Harvey, C. R. (1999). Conditioning variables and the cross section of stock returns. *The Journal of Finance*, *54*(4), 1325-1360.

Ferson, W. E., & Lin, J. (2014). Alpha and performance measurement: the effects of investor disagreement and heterogeneity. *The Journal of Finance*, 69(4), 1565-1596.

Frazzini, A., & Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, *111*(1), 1-25.

Frazzini, A., Israel, R., & Moskowitz, T. J. (2012). Trading costs of asset pricing anomalies. *Fama-Miller Working Paper*, 14-05.

French, K.R. (2008). Presidential address: The cost of active investing. *The Journal of Finance,* 63(4), 1537-1573.

Gibbons, M. R., Ross, S. A., & Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica: Journal of the Econometric Society*, 1121-1152.

Green, J., Hand, J. R., & Zhang, F. (2014). The remarkable multidimensionality in the cross-section of expected US stock returns. *SSRN working paper No. 2262374*.

Green,J., Hand, J. R., & Zhang, X.F. (2013). The supraview of return predictive signals. *Review of Accounting Studies*, 18(3), 692-730.

Hansen, L. P., & Jagannathan, R. (1991). Implications of security market data for models of dynamic economies. *Journal of Political Economy*, *99*(2), 225-262.

Hansen, L. P., & Jagannathan, R. (1991). Restrictions on intertemporal marginal rates of substitution implied by asset returns. *Journal of Political Economy*, *99*(2), 225-62.

Harvey, C. R., Liu, Y., & Zhu, H. (2015). ... And the cross-section of expected returns. *Review of Financial Studies*, 29(1), 5-68.

Idzorek, T. M., & Kowara, M. (2013). Factor-Based Asset Allocation vs. Asset-Class-Based Asset Allocation. *Financial Analysts Journal*, 69(3), 19-29.

Jacobs, H (2015). What explains the dynamics of 100 anomalies?. *Journal of Banking & Finance, 57, 65- 85.*

Jagannathan, R., & Wang, Z. (1996). The conditional CAPM and the cross‐section of expected returns. *The Journal of Finance*, *51*(1), 3-53.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, *48*(1), 65-91.

Jegadeesh, N., & Titman, S. (2001). Profitability of momentum strategies: An evaluation of alternative explanations. *The Journal of Finance*, *56*(2), 699-720.

Jobson, J. D., & Korkie, B. M. (1981). Performance hypothesis testing with the Sharpe and Treynor measures. *The Journal of Finance*, *36*(4), 889-908.

Kandel, S., & Stambaugh, R.F. (1987). On correlations and inferences about mean-variance efficiency. *Journal of Financial Economics,* 18(1), 61-90.
Karolyi, G. A. (1992). Predicting risk: some new generalizations. *Management Science*, *38*(1), 57-74.
Kozak, S., Nagel, S., & Santosh, S. (2017). Interpreting factor models. *Journal of Finance, Forthcoming.*
Lakonishok, Josef, and Alan C. Shapiro, (1986). Systematic risk, total risk and size as determinants of stock market returns, *Journal of Banking and Finance* 10, 115-132.
Ledoit, O., & Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. Journal of Empirical Finance, 15(5), 850-859.
Lewellen, J. (2015). The Cross-section of Expected Stock Returns. *Critical Finance Review*, *4*(1), 1-44.
Lewellen, J., & Nagel, S. (2006). The conditional CAPM does not explain asset-pricing anomalies. *Journal of Financial Economics*, *82*(2), 289-314.
McLean, R. D., & Pontiff, J. (2016). Does academic research destroy stock return predictability?. *The Journal of Finance*, *71*(1), 5-32.
McLean, R.D., Pontiff, J., & Watanabe, A. (2009). Share issuance and the cross-sectional returns: International evidence. *Journal of Financial Economics,* 94(1), 1-17.
Memmel, C. (2003). Performance hypothesis testing with the Sharpe ratio. *Finance Letters*, *1*(1).
Miller, E. M. (1977). Risk, uncertainty, and divergence of opinion. *The Journal of Finance*, *32*(4), 1151-1168.
Novy-Marx, R., & Velikov, M. (2016). A taxonomy of anomalies and their trading costs. *Review of Financial Studies*, *29*(1), 104-147.
Opdyke, J. D. J. (2007). Comparing Sharpe ratios: so where are the p-values?. *Journal of Asset Management*, *8*(5), 308-336.
Pontiff, J., & Woodgate, A. (2008). Share issuance and cross‐sectional returns. *The Journal of Finance*, *63*(2), 921-945.
Renneboog, L., Ter Horst, J., & Zhang, C. (2008). Socially responsible investments: Institutional aspects, performance, and investor behavior. *Journal of Banking & Finance*, *32*(9), 1723-1742.
Roll, R. (1977). A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory. *Journal of Financial Economics*, 4(2), 129-176.
Rosenberg, B., Reid, K., & Lanstein, R. (1985). Persuasive evidence of market inefficiency. *The Journal of Portfolio Management*, 11(3), 9-16.
Shanken, J. (1987). Multivariate proxies and asset pricing relations: Living with the Roll critique. *Journal of Financial Economics,* 18(1), 91-110.
Shiller, R. J. (1982, January). Consumption, asset markets and macroeconomic fluctuations. In *Carnegie-Rochester Conference Series on Public Policy* (Vol. 17, pp. 203-238). North-Holland.
Shleifer, A., & Vishny, R. W. (1997). The limits of arbitrage. *The Journal of Finance,* 52(1), 35-55.
Sloan, R. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings?(Digest summary). *Accounting review*, *71*(3), 289-315.
Stambaugh, R. F. (1982). On the exclusion of assets from tests of the two-parameter model: A sensitivity analysis. *Journal of Financial Economics*,*10*(3), 237-268.
Stambaugh, R. F., Yu, J., & Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, 104(2), 288-302.
Stattman, D. (1980). Book values and stock returns. *The Chicago MBA: A journal of selected papers*, *4*(1), 25-45.
Titman, S., Wei, K. J., & Xie, F. (2004). Capital Investments and Stock Returns. *Journal of Financial and Quantitative Analysis*, 677-700.
Vasicek, O. A. (1973). A note on using cross-sectional information in Bayesian estimation of security betas. *The Journal of Finance*, *28*(5), 1233-1239.
White, H., (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, 817–838.

Appendix A – Hansen-Jagganathan bound

This section derives the Hansen-Jagganathan bound formally and discusses its intuition for rational testing:

$$\frac{\sigma_t(M_{t+1})}{E_t(M_{t+1})} \geq \frac{E_t(R_{i,t+1}^{ex})}{\sigma_t(R_{i,t+1}^{ex})}$$

Where E_t is conditional expectation on today's information, σ_t is the standard deviation, M_{t+1} is the stochastic discount factor and $R_{i,t+1}^{ex}$ is excess returns over the risk-free rate for asset i. The Hansen-Jagganathan (HJ) bound states that the portfolio of assets with the highest Sharpe ratio puts a lower bound on the volatility of the SDF (Shiller, 1982; Hansen & Jagannathan, 1991; Campbell, 2000). Decomposing a portfolio's standard deviation gives further insights.

$$\frac{\sigma_t(M_{t+1})}{E_t(M_{t+1})} \geq \frac{E_t(R_{i,t+1}^{ex})}{\sqrt{\sum_{k=1}^{N}\sum_{j=1}^{N} w_k\, w_j \sigma_k \sigma_j p_{kj}}}$$

Where σ_k and σ_j is the standard deviation of comprising assets and p_{kj} is their correlation. Holding all else constant, as the correlation among premiums falls, the maximum Sharpe ratio of the combined strategy rises, which increases the difficulty of fitting a discount factor in the HJ bounds.

Furthermore, if we assume absence of correlation between assets in an equally weighted portfolio, the bound reduces to:

$$\frac{\sigma_t(M_{t+1})}{E_t(M_{t+1})} \geq \frac{E_t(R_{i,t+1}^{ex})}{\sqrt{\frac{1}{N}[\sigma_j']}}$$

Where σ_j' is the average standard deviation of comprising portfolio assets. As N goes to infinity the variance of the portfolio goes to zero. The HJ bound gives a good intuition as to why the existence of many uncorrelated positive excess return bets makes rational pricing more difficult.

Appendix B - Fama-French benchmarks

Benchmarking against anomalies that comprise the three-factor (Fama & French, 1993) or the five-factor model (Fama & French, 2015), also gives economically large and statistically significant alpha estimates that easily pass the new data mining adjusted t-statistics hurdles (Table 7). Even profitability passes the higher t-statistic hurdle of 3 when benchmarked against the FF3 premiums. In fact, the average alpha t-statistic of non-benchmark anomalies *increases* considerably when size and value are used as benchmarks (average alpha t-statistic of 5.35) instead of the term structure and default premiums (average alpha t-statistics of 4.34). This is intuitive as second-generation anomalies were developed as violations of the Fama-French three factor model. Overall, the result suggests that the use of a passive benchmark is not particularly lenient when it comes to assessing anomaly significance. Above and beyond that, it shows that non-benchmark anomalies are not linear combinations of benchmark anomalies.

Conducting this specific robustness check has its merits. Nevertheless, there are several notable drawbacks specific to this approach. First and foremost, anomaly alpha when benchmarked against other anomalies is meaningless if investors do not already hold the benchmark anomalies to begin with. In other words, if investors do not already hold size, value, investment, and profitability in their portfolio, benchmarking against them does not make much sense. Second, assessment with anomaly benchmarks is highly sensitive to the arbitrary choice of initial anomalies to be included in the benchmark. For example, if quality is first included in the benchmark, then the alpha of size becomes significant (Asness, Frazzini & Pedersen, 2015). Stated alternatively, the order of adding anomalies to the benchmark can matter. Finally, anomaly benchmarking is a data reduction technique that is particularly exposed to the benchmark ambiguity problem. Having mentioned these caveats, the academic tradition of showing anomaly alpha with respect to other anomalies, as a *complement* to the regressions with passive proxies, is a good starting point for assessing between anomaly substitutability.

	FF3				FF5					
	α	β	υ	ι	ά	ß	ύ	ί	θ	ζ
RMW	0.51%	-0.26	-0.38	-0.10						
CMA	0.38%	-0.19	0.01	0.34						
WML	1.60%	-0.26	0.05	-0.34	1.34%	-0.13	0.12	-0.46	0.19	0.42
IVOL	1.08%	-0.91	-0.88	0.28	0.60%	-0.67	-0.62	0.23	0.69	0.32
QUAL	0.90%	-0.42	-0.43	-0.25	0.56%	-0.24	-0.22	-0.25	0.56	0.15
BAB	0.77%	-0.06	0.00	0.18	0.52%	0.07	0.12	0.14	0.34	0.21
AC	0.44%	-0.07	-0.10	0.06	0.45%	-0.07	-0.14	0.01	-0.10	0.12
NI	0.59%	-0.22	-0.30	0.09	0.35%	-0.10	-0.23	-0.01	0.21	0.36
μ	*0.79%*	*-0.30*	*-0.25*	*0.03*	*0.64%*	*-0.19*	*-0.16*	*-0.06*	*0.31*	*0.26*

	t-stats				*t-stats*					
	t(α)	***t(β)***	***t(υ)***	***t(ι)***	***t(ά)***	***t(ß)***	***t(ύ)***	***t(ί)***	***t(θ)***	***t(ζ)***
RMW	***4.11***	*-6.41*	*-10.25*	*-2.15*						
CMA	***3.56***	*-6.74*	*0.24*	*10.65*						
WML	***6.09***	*-3.08*	*0.48*	*-2.93*	***4.58***	*-1.48*	*1.18*	*-3.32*	*1.62*	*2.57*
IVOL	***5.58***	*-13.3*	*-17.65*	*3.42*	***3.45***	*-11.22*	*-11.45*	*2.47*	*7.87*	*3.61*
QUAL	***8.15***	*-12.37*	*-16.97*	*-8.06*	***6.48***	*-9.82*	*-9.38*	*-10.1*	*16.4*	*3.94*
BAB	***5.91***	*-1.24*	*-0.06*	*3.82*	***4.10***	*1.50*	*3.23*	*2.46*	*7.47*	*3.18*
AC	***3.87***	*-2.34*	*-3.12*	*2.12*	***3.83***	*-2.36*	*-3.64*	*0.26*	*-1.77*	*2.34*
NI	***5.50***	*-7.31*	*-13.06*	*3.04*	***3.44***	*-3.38*	*-8.55*	*-0.43*	*5.75*	*8.45*
μ	***5.35***	*-6.60*	*-7.55*	*1.24*	***4.31***	*-4.46*	*-4.77*	*-1.44*	*6.23*	*4.02*

Table 7: Benchmark regressions. Uses monthly US data from 07/1963 until 12/2014. MKT is the equity premium, GOV is the term structure premium, CORP is the default premium, SMB is size, BTM is value, RWM is profitability, CMA is investment, WML is momentum, IVOL is idiosyncratic volatility, QUAL is quality, BAB is betting against beta, AC is accruals and NI is net share issuance. Reported are the alphas and loadings from constant coefficient unconditional time-series regressions:

(FF3) $$R_{i,t} = \alpha_i + \beta_i MKT_t + \upsilon_i SBM_t + \iota_i HML_t + \varepsilon_{i,t}$$

(FF5) $$R_{i,t} = \acute{\alpha}_i + \text{ß}_i MKT_t + \acute{\upsilon}_i SMB_t + \acute{\iota}_i HML_t + \theta_i RMW_t + \zeta_i CMA_t + \acute{\varepsilon}_{i,t}$$

where $R_{i,t}$ is the return of strategy i in month t. The regressions are constant coefficient unconditional. T-statistics are computed using the heteroskedasticity consistent standard errors of White (1980).

Appendix C – Long sample

As an additional robustness check, the paper investigates performance ordering in an extended sample using the three most well-known anomalies; Table 8 contains the results for a robustness check in a long sample starting from 01/1927 that covers only size, value and momentum as active investment options. Results in this long sample support the overall conclusions. The Sharpe ratio of the balanced portfolio is both economically and statistically larger than the Sharpe ratio of the traditional portfolio. The improvement in performance is corroborated across the Sortino ratio, recession t-statistic and maximum drawdown measures. Overall, the robustness check suggests that the main findings are not specific to the chosen anomalies and time frame.

	TP	FP	BP
R	0.30%	0.40%	0.40%
t-stat	*5.19*	*7.71*	*8.75*
R Recession	0.05%	0.20%	0.13%
t-stat	*0.22*	*1.43*	*1.11*
SD	2.17%	1.82%	1.45%
SD Recessions	3.27%	2.05%	1.64%
SD Expansions	1.78%	1.75%	1.38%
M2	10.40%	15.40%	17.50%
Skewness	0.22	0.6	0.49
t-stat	*2.91*	*7.91*	*6.51*
Excess Kurtosis	4.98	9.25	8.71
t-stat	*33.02*	*61.36*	*57.75*
JB test	1086	3787	3342
Max DR	41%	24%	**16%**
Target DD	1.40%	1.10%	0.80%
Sortino Ratio	0.25	0.41	**0.46**
Sharpe Ratio	0.16	0.24	**0.27**
LW Bootstrap p-val (TP)	0.005	0.104	**0.00**
TP Beta	1	0.04	0.52

Table 8: Long sample. Displays the performance of portfolios in a long sample from 01/1927 until 12/2014. A traditional portfolio (TP) is an EW combination of the market, government bond and corporate bond premiums. The factor portfolio is made as an EW portfolio of the size, value, and momentum premiums. The premium portfolio is a per premium equally weighted portfolio containing all factor and traditional premiums. The premium portfolio in the long sample is equivalent to the balanced portfolio. TP Beta is the beta from a time series regression on the traditional portfolio. The p-value for the Ledoit-Wolf Bootstrap TP and BP test is calculated with respect to the traditional portfolio (TP) (and the equity premium for the traditional portfolio itself). The target for the Sortino ratio and downside deviation is zero. M2 is the annualized Modigliani & Modigliani measure which volatility matches strategies to a benchmark (the equity premium). JB is the Jarque-Bera test for normality. The critical values for the test are 10% (4.38), 5% (5.88), 1% (10.53). Maximum drawdown (Max DR) is the maximum percentage drop from a peak. T-statistics for the mean are computed using the heteroskedasticity consistent standard errors of White (1980).

Appendix D – Supplementary tables and figures

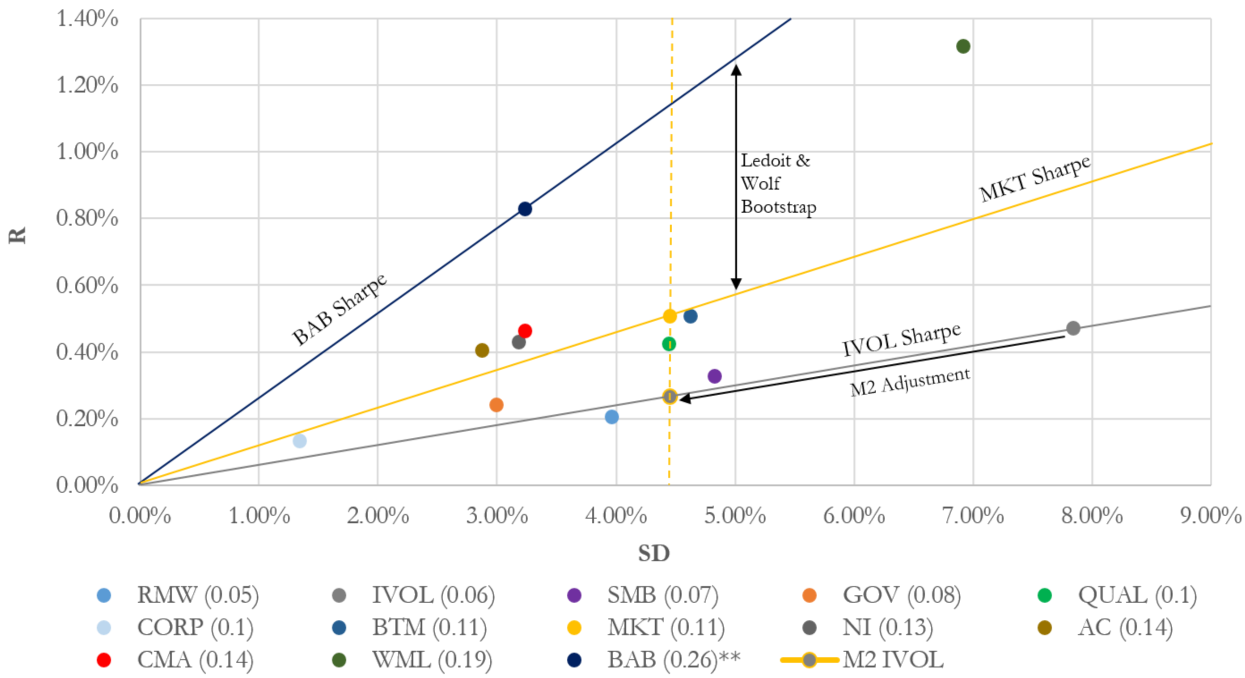

Figure 9: **Sharpe Ratios.** Displays the performance of portfolios in return/standard deviation space during the sample period spanning 07/1963-12/2014. Realized Sharpe Ratios are displayed in brackets below the figure. MKT is the equity premium, GOV is the term structure premium, CORP is the default premium, SMB is the size, BTM is value, RWM is profitability, CMA is investment, WML is momentum, IVOL is idiosyncratic volatility, QUAL is quality, BAB is betting against beta, AC is accruals and NI is net share issuance. M2 is the Modigliani & Modigliani measure which volatility matches strategies to a benchmark (the equity premium in this case). Statistical significance of differences in Sharpe ratios is calculated using the Ledoit-Wolf bootstrap test (with the equity premium Sharpe Ratio). Asterisk denote statistical significance at the 5% (**) level.

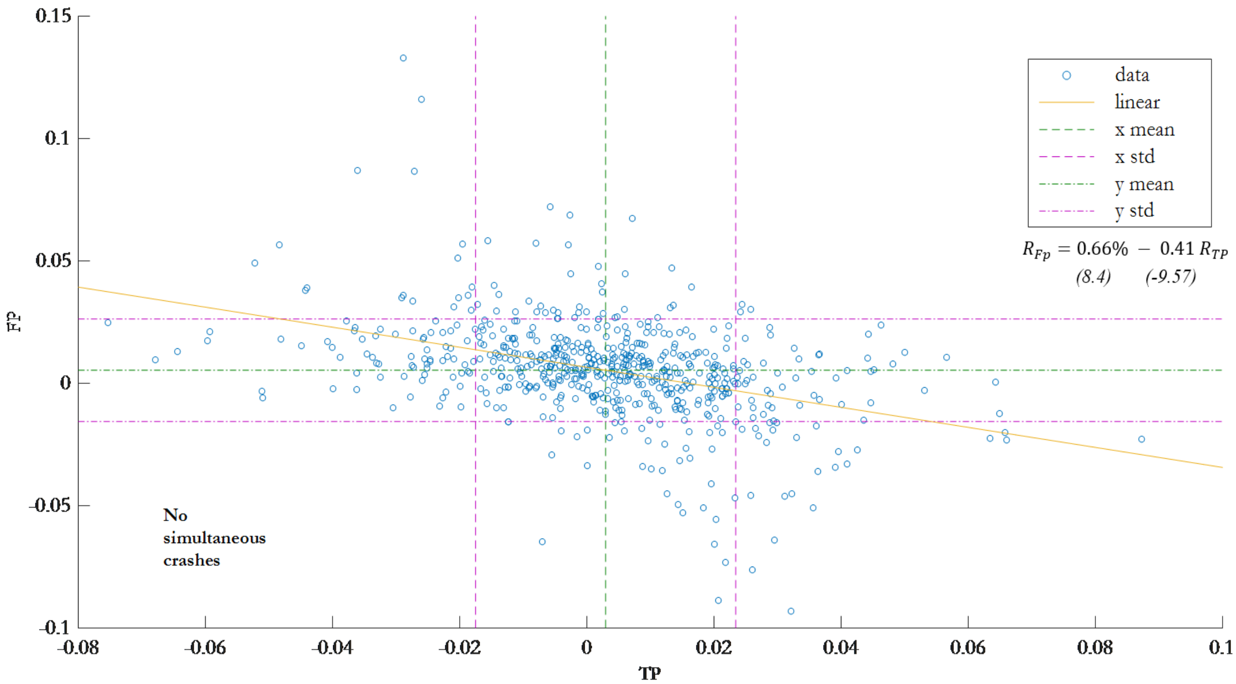

Figure 10: Absence of simultaneous crashes. Displays the realization pairs for the traditional and factor portfolio. Loadings and alphas are calculated using constant coefficient unconditional time-series regressions:

$$R_{FP,t} = \acute{\alpha}_i + \lambda_i R_{TP,t} + \acute{\varepsilon}_{it}.$$

The traditional portfolio (TP) is constructed as an EW portfolio of equity (MKT), term structure (GOV) and default (CORP) premiums. The factor portfolio (FP) is constructed as an equally weighted portfolio of size, value, profitability, investment, momentum, idiosyncratic volatility, quality, betting against beta, accruals and net issuance. Standard errors are calculated using the heteroskedasticity consistent standard errors of White (1980).