Journal Of Investment Management, Vol. 17, No. 3, (2019), pp. 65–81

www.joim.com

BILL GROSS' ALPHA: THE KING VERSUS THE ORACLE

Aaron Brown[a,*] *and Richard Dewey*[b,†]

We set out to investigate whether "Bond King" Bill Gross demonstrated alpha (excess average return after adjusting for market exposures) over his career, in the spirit of earlier papers asking the same question of "Oracle of Omaha," Warren Buffett. The journey turned out to be more interesting than the destination. We do find, contrary to previous research, that Gross demonstrated alpha at conventional levels of statistical significance. But we also find that result depends less on the historical record than on whether we take the perspective of academics interested in market efficiency, investors picking a fund or someone (say a potential employer) asking whether a manager has skill or is throwing darts to pick positions. These are often thought to be overlapping or even identical questions. That is not completely unreasonable in equity markets, but in fixed income these are distinct. We also find quantitative differences, mainly that fixed-income securities have much higher correlations with each other than equities, make alpha 4.5 times as hard to measure for Gross than Buffett. We do not think our results will have much practical effect on attitudes toward Gross as an investor, but we hope they will advance understanding of what alpha means and appropriate ways to estimate it.

1 Introduction

Superstar bond portfolio manager Bill Gross announced his retirement last week. From 1987 to 2014, his PIMCO Total Return fund generated 1.33% per year of alpha versus the Barclays US Credit Index, with a t-statistic of 3.76. For many years his fund was the largest bond fund in the world, and was generally considered to be the most successful.

This track record inspired us to take a closer quantitative look along the lines of Frazzini, Kabiller and Pedersen's (2018) *Buffett's Alpha* (FKP). Gross, like Buffett, often publicly discussed what he perceives as the drivers of his returns. At

[a]The author was employed at AQR Capital Management during the period two key papers cited here were written, and he discussed them with the authors prior to publication.
[b]The author worked at PIMCO from Jan 2015 through the Dec 2016. He was not employed at PIMCO during any of the periods studied in this paper.
*New York University, USA. E-mail: acb22@nyu.edu
†Royal Bridge Capital, USA. E-mail: rich@royalbridge-cap.com

the Morningstar Conference in 2014 and in a 2005 paper titled *"Consistent Alpha Generation Through Structure"* Gross highlighted three factors behind his returns: more credit risk than his benchmark, more 5-year and less 30-year exposure, and long mortgages and other securities with negative convexity.

We present five main findings:

1. We confirm that those three factors, plus one for the general level of interest rates, explain 89% of the variance in Gross' monthly return over the 27-year period. We further estimate that Gross outperformed a passive factor portfolio by 0.84% per year, which is significant at the 5% level. Gross' compounded annual return over the period was 7.52%, versus 6.44% for the Barclay's Aggregate US Index. So we find that most of his 1.08% annual outperformance of the index was alpha.
2. The FKP paper mentioned above considered one of the best-known track records in the equity asset class, Warren Buffett's. We compliment this work by examining one of the best-known track records in the fixed-income asset class. Fixed-income investing offers a different set of challenges and opportunities than equity. We offer a novel discussion on the concept of manager alpha including important qualitative and quantitative differences in the concept of alpha with Gross versus Buffett.
3. The main qualitative difference is that Gross exploited well-known sources of risk and potentially excess return in the fixed-income market, exposures that investors rationally demand additional yield to accept. Buffett's performance, for the most part, correlates with factors uncovered long after he began investing and were still not accepted as fully as factors like credit risk or mortgage prepayment risk. Moreover, Buffett's factors probably result from behavioral biases and institutional constraints rather than rational investor preferences.
4. The main statistical difference is the much higher r^2 value in Gross' regression versus Buffett's (about 0.9 versus 0.3) makes the alpha significance estimate 4.5 times as sensitive to the observed returns on the factor portfolios. Since it is nearly impossible to estimate expected returns — there is considerable debate about the level of the equity premium even with 150 years or more of data–this makes it important to select factors that conform as closely as possible to what Gross actually did, rather than factors that merely have a high return correlation to Gross's results. The closer the factors conform to Gross' practice, the better the chance that any deviations in factor performance from expectation over the period are reflected equally in both Gross's actual results and the factor portfolio results.
5. Gross earned essentially all of his alpha in favorable markets for his factors and had a significantly negative timing ability in the sense that his factor exposures were greater in months the factor had negative returns than in months the factor had positive return. This latter feature could be unfortunate timing decisions or negative convexity in the factor exposures. We discuss whether this can shed light on the source of Gross's alpha, specifically whether it relates to preferential access to new issues and leverage.

This work has implications for market efficiency, portfolio construction and manager selection. Does Gross' performance offer evidence against efficiency in fixed-income markets? Is the track record better than what we would expect given its factor exposures? And, how should an investor view the data?

The remainder of this paper is organized as follows: Section 2 offers a brief literature review.

Section 3 describes the data. Section 4 describes the methodology. Section 5 offers a discussion on manager alpha. Section 6 details our results and is followed by our conclusions.

2 Literature review

There is an extensive literature that examines fixed-income markets for information in the yield curve and return predictability. Fama and Bliss (1988) and Shiller and Campbell (1991) run regressions on yield spreads and find return predictability. Ilmanen (1995) studies relative risk aversion, time-varying risk premium (historical bond market beta), real bond yields and the term spread and finds significant return predictability. Cochrane and Piazzesi (2005) find a forward factor that predicts returns at statistically and economically significant levels. Litterman and Scheinkman (1991) identify interest rate sensitivity, term structure sensitivity and volatility as drivers of return.

There is also a robust literature that seeks to identify common return factors in equity markets. This literature goes back at least to Fama and French (1992) who documented three equity market factors (market, small cap and value) and two fixed-income factors (term premium and credit premium). Following their work, factor research focused mostly on equity markets with Asness (1995) and Carhart (1997) documenting cross-sectional momentum, Asness *et al.* (2014) showing a quality factor and Frazzini and Pedersen (2014) uncovering the low beta premium to name just a few of the better established equity market factors.

These two distinct strands of literature notwithstanding, there has been relatively little work in looking at fixed-income markets through the lens of cross-sectional factors in a similar spirit to equity markets, although recently some research has emerged. Burham (2016) looks at the low beta anomaly in fixed-income markets and finds that overweighting lower beta fixed-income instruments, which typically tend to have shorter maturities is a persistent source of alpha. Several AQR practitioner papers find factors including value, momentum and carry in credit and government bonds markets to be sources of alpha in the fixed-income markets. Brooks and Moskowitz (2017) look at the cross-section of returns in government bond markets and find that much of the aforementioned literature looking at slope and curvature of the yield curve can actually be explained by factors such as value, momentum and carry. A recent AQR paper (2017) examines active fixed-income portfolio management and argues that a significant portion of the returns are down to a passive bet on high-yield bonds.

Our work is related to this recent literature as we attempt to decompose the returns of PIMCO's Total Return fund to determine if cross-sectional overweighting of factors such as low beta in bonds, increased tilts toward negative convexity assets, and higher credit exposure can explain the fund's historical outperformance. Our work is related to the literature on performance attribution of investment managers. Brooks *et al.* (2019) build on FKP (2018) and examine the track records of superstar performers including Gross, and attempt to decompose the returns according to common factors. This paper is the most similar to ours and we compare results below. Our research is also related to that of Lo and Lee (2014) and Jurek and Stafford (2015), both of whom perform attribution on hedge fund returns and find that common factors drive a significant amount of performance.

3 Data

Our analysis is confined to historical returns, whereas FKP hand collected extensive data on actual holdings. Hand collecting data for Gross

Table 1 Data used in this study along with their sources. All data except the 1-month treasury was sourced from Bloomberg and only the low beta (5s/30s factor required calculation on our part.

Factor	Time period	Source	Annualized compound return	Annualized standard deviation	Annualized Sharpe ratio
Overview of data					
PIMCO TR	1987–2014	Bloomberg	4.29%	4.23%	0.20
BarCap US Credit Index	1987–2014	Bloomberg	3.92%	5.15%	0.09
BarCap MBS	1987–2014	Bloomberg	3.32%	3.17%	−0.04
5s/30s	1987–2014	Bloomberg	1.42%	2.41%	−0.84
10-yr TSY	1987–2014	Bloomberg	3.96%	6.21%	0.08
1M Treasury	1987–2014	Ken French	3.45%	0.71%	0

proved too difficult, because the public portfolio records omit necessary details about many of his over-the-counter swaps, options and deferred payment arrangements, which were highly leveraged so even small errors in approximation of individual positions amplify to large errors in portfolio exposures.

The Total Return Fund was launched in May 1987. We begin with the month of June 1987 as the first full month of operation. We end in September 2014 when Gross resigned from managing the fund. Gross' performance since leaving PIMCO has been less impressive. We do not include that data as the institutional setting in which it was produced was quite different.

PIMCO TR was benchmarked to Barclays Aggregate US Index. However we choose instead to state our main results versus the broader Barclays US Credit Index as it has higher correlation to Gross' returns and its holdings are a closer match as well (the "US" means US dollar denominated only, so unlike the Barclays Agg Index, the US Credit Index holds bonds from global issuers, as did PIMCO TR). We also make use of the Merrill Lynch 10-Year Treasury Index and the Barclays US MBS Index.

Fixed-Income indices are inherently more complex than equity indices. Bonds change character over time as their maturity dates get closer, and bond indices have time-varying average coupons, average credit spread, issuer composition and other key financial attributes. Price data are not of the same quality or standard as in the equity literature. The fixed-income market, until recently, was an inter-dealer broker market and reliable, consolidated prices in standardized databases for our full sample period do not exist to the best of our knowledge.[1] Since bonds trade less frequently than equities and at larger spreads, price series necessarily contain a mix of opinion and observation of arms-length transactions. This issue is particularly relevant because one common reason offered for Gross' superior returns is access to new issues and derivatives on favorable terms as well as preferred financing arrangements. If this is correct, the prices available to PIMCO TR may not correspond with the prices used to compute index values.

We use interest rates futures data to estimate the returns on a zero-duration portfolio long 5-year treasuries and short 30-year, and the 1-month treasury yield as a risk-free rate.

All data except the risk-free rate are from Bloomberg. Only the 5s/30s factor requires

Table 2 Correlation coefficients between monthly returns of the factors from 1987 to 2014.

	10-Year Treasury	Credit	MBS	5s/30s
Factor correlation matrix				
10-Year Treasury	1.00	0.77	0.87	0.68
Credit		1.00	0.78	0.45
MBS			1.00	0.65
5s/30s				1.00

calculation by us. We obtained the 1-month treasury rate from Ibbotson via Ken French's website.

Our first three factors have high correlations because they are all long fixed-income instruments that respond mainly to interest rates. The correlations of each factor can be found in Table 2. Our 5s/30s factor is calibrated to zero-duration, but only under the assumption of parallel shifts in the yield curve. In fact, since the 5-year yield is much more volatile than the 30-year yield, when rates fall the yield curve tends to steepen (the 5-year rate falls more than the 30-year rate), so the steepener has positive returns when bonds in general have positive returns. The correlations with 5s/30s are lower than the correlations among the other three factors, because the indirect yield curve shape effect is not as strong as the direct interest rate exposure effect.

Table 3 shows regressions performed for the four main factors that will be explored in this paper. One takeaway from the factor regressions is that Credit has a negative conditional exposure to 5s/30s. We found this surprising, since steeper yield curves are usually associated with better credit conditions, while flat or inverted yield curves are taken as signs of impending recession. However, at least over this period of time, conditional on the level of interest rates (and also on mortgage security returns) a steepening yield curve was associated with negative returns on credit, and a flattening yield curve was associated with positive returns on credit.

Another takeaway is that 5s/30s and Credit have more independent volatility from the other three factors than the 10-Year Treasury and MBS do. Finally, within the set of four factors, MBS has significant positive alpha, the 10-Year Treasury

Table 3 Regression results over our entire sample period from 1987 to 2014. These are four separate multiple regressions with the dependent variables listed in the first column. Each one is regressed on the other three factors. t-Statistics are shown below coefficients in parentheses.

	10-Year Treasury	Credit	MBS	5s/30s	Annualized alpha	R^2
Factor Regressions						
10-Year Treasury		0.32	1.02	0.56	−1.37%	0.80
(t-stat)		(6.7)	(11.2)	(6.7)	(−2.5)	
Credit	0.38		0.81	−0.40	0.31%	0.66
(t-stat)	(6.7)		(7.5)	(−4.2)	(0.50)	
MBS	0.27	0.18		0.20	1.19%	0.80
(t-stat)	(11.2)	(7.5)		(4.5)	(4.2)	
5s/30s	0.21	−0.13	2.9		0.10%	0.50
(t-stat)	(6.7)	(−4.2)	(4.5)		(0.3)	

has significant negative alpha while the alphas of Credit and 5s/30s cannot be distinguished from zero at the 5% level.

4 Methodology and factor construction

4.1 Interest rates, credit and convexity

The usual practice in the factor literature is to use a single long-only market factor minus the risk-free rate, and zero-investment long–short portfolios for all other factors. Each of these long–short factors isolates a single economic return contribution. For reasons discussed below we chose a different approach. We elect to use long-only market factors for three of our factors: Interest rates (Merrill Lynch 10-Year Treasury Index), credit-sensitive bonds (Barclays US Credit) and mortgages (Barclays US MBS).

4.2 Yield curve steepener

Gaining exposure to the duration neutral 5s/30s curve position requires an investor to buy a 5-yr Treasury note and sell short a duration equivalent 30-yr Treasury bond, or to take the equivalent position using futures.

The carry of a government bond can be decomposed into the coupon income and the rolldown. Rolldown describes the effect of time passing and bonds increasing in value as they roll down the curve to a lower yield and higher price. The interest rate curve must be upward sloping for this to be positive. This has been the case more often than not and Mr. Gross has argued that in a highly leveraged economy that this upward sloping curve must prevail the majority of the time.

The duration-neutral bond steepening position is constructed as follows:

$$\text{Weight}_t^{\text{STEEPENER}} = \begin{cases} 5yrBond_t \times Dur_{5yr} \\ 30yrBond_t \times Dur_{5yr/30Yr} \end{cases}$$

The idea behind using a 5's/30's steepener position is not to make trades around the curvature of the yield curve, but to benefit from carry. Typically the 30-year bond carries a higher coupon than the 5-year bond, but an equal duration levered 5-year portfolio earns more net coupon.

Moreover the 5-year usually earns more rolldown because the curve is steeper at the 5-year point, i.e. the difference between the 5-year and 4-year yield is greater than the difference between the 30-year and the 29-year.

Historically, the coupon and rolldown effects make a levered 5-year position usually return more than an unlevered 30-year position with the same duration.

5 Measuring manager alpha

5.1 Meaning of alpha

Before we discuss our results, we define alpha and discuss what it means. The financial concept of alpha was developed in the context of a single-factor pricing model. A portfolio manager who could deliver consistent positive alpha was a skilled investor, a good choice for investors to hire and evidence against market efficiency. However, in a multi-factor context, skill, good choice and market efficiency are separate questions; and in this respect Buffett is different from Gross.

In a single-factor model, the market is assumed to carry a return premium because investors dislike the risk. Since it is the risk of the entire market it cannot be diversified away (cross-sectionally, that is, it can be diversified over time), and it can only be hedged by selling it to someone else. An investor earning higher-than-market returns by taking more market exposure is not demonstrating skill, helping her investors nor providing evidence against market efficiency.

With multi-factor models some factors are usually explained as things investors dislike, just like market risk, while others may be explained as institutional constraint or persistent behavioral anomaly. Not everyone agrees with the explanations for each factor, or even whether both kinds of factors exist. But most researchers would agree that high book-value-to-price stocks, for example, probably owe their return premium to institutional and behavioral tendencies. There seems small plausible reason for investors to care whether their portfolios are correlated to the value factor. Exposure to the value factor is not the same as owning stocks with high book-value-to-price ratios as there are other ways to obtain it, and not all high book-value-to-price ratio stocks are correlated to the factor; so even if an investor cares about the financial valuation ratios of the portfolio, that is not the same as caring about exposure to the value factor. On the other hand rational leverage aversion, because leverage can inflict hard-to-manage tail risk at the most painful times, could explain a low beta factor. The difference between these two types of factors is important when interpreting alpha.

FKP decomposed Warren Buffett's returns into exposures to the stock market, value, quality and low beta, after which his residual alpha was positive (6.3% per year), but with a t-statistic below 2 (1.58). Previous work using different factors had found t-statistics on the residual alpha greater than 2, which was evidence against market efficiency. But the difference does not matter to an investor considering investing with Buffett by buying Berkshire Hathaway stock. Buffett's return was better than the market, and the additional exposures to two of the three factors, and possibly all three, does not detract from value of those excess returns to investors. Investors might care for some indirect reasons. For example, if Buffett's returns are correlated to value and so are another manager's, the investor might get less diversification than she had hoped for. But this depends on the total correlation among managers. Two managers can both have positive correlations to the same factors, without being correlated to each other, especially for the relatively small correlations found in FKP. An investor might also care if someone else is selling the same collection of factor exposures cheaper. Then the question of whether Buffett's alpha is statistically significant or not matters. And even if an investor cares about value exposure for some indirect reasons, there is no reason to assume the reduction in her utility is equal to the excess return on the factor over the period of study. If we want a theory of how investors should value Buffett's performance, we need to specify investor utility to factor exposures exogenously. Therefore, FKP does not have much to say about whether or not investing with Warren Buffett is a good idea.

There is a third reason people look to alpha, to determine if a manager has skill. However, as FKP state: "[I]t cannot be emphasized enough that explaining Buffett's performance with the benefit of hindsight does not diminish his outstanding accomplishment. He decided to invest based on these principles half a century ago. He found a way to apply leverage. Finally, he managed to stick to his principles and continue operating at high risk even after experiencing some ups and downs that have caused many other investors to rethink and retreat from their original strategies." We would add another important skill, communicating with investors so that they did not rethink and retreat.

5.2 Gross and market efficiency

The case of Gross is different in all three respects: evidence for market efficiency, importance of alpha for investors choosing managers and alpha as a measure of skill. Moreover, there is a key quantitative difference.

Because the market factor we use to explain Gross' return is restricted to the US dollar denominated, fixed-rate, investment grade fixed-income market, statistically significant alpha is no evidence against general market efficiency. An equity index fund will have statistically significant alpha against this index. The market factor in FKP was the CRSP Value-Weighted US stocks. While that is also not the complete portfolio of all global investible assets, it has a reasonably high correlation to that idealized factor, because equity risk dominates total financial risk and global equities are correlated to US equities.

If Gross has significant alpha with respect to a broad fixed-income index and other factors, using similar securities to those in the market, then that could be taken as evidence that the fixed-income market is not efficient internally. We might ask why all fixed-income investors did not do what Gross did. However, it is well known that fixed-income investors do not all share the same measures of risk.

An investor with long-term nominal liabilities thinks short-term bonds are risky and long-term ones are safe, the opposite of someone looking at daily mark-to-market volatility. Investors have differing natural exposures to inflation, real rates, currencies and other factors. While the same thing might be said of equity investors, it is less obvious, and the volatility of these natural exposures is smaller in relation to the volatility of equities. Generally speaking, investors want high returns from their equity funds, correlations to economic variables other than the overall market are secondary considerations; while many fixed-income investors have primary concern with economic sensitivities.

For these reasons, we claim our results cannot be used to argue for or against market efficiency, either in a broad or narrow sense.

5.3 Was Gross a good choice for investors?

Two of the factors we will use to explain PIMCO TR return above benchmark are credit risk and short volatility. Most fixed-income investors care about these exposures. One reason is that these factors usually give small positive returns, and inflict occasional large losses, typically at the most painful times, a return pattern investors dislike. Another reason is that it is hard to evaluate the amount of left-tail risk in these strategies.

The third factor we use to explain Gross' returns versus the market is the analog to the low beta factor in equities, 5s/30s or overweighting short maturities and underweighting long ones. The most likely reason for this strategy to carry a return premium is leverage aversion among investors. Some investors might care that Gross achieved some of his returns by using leverage, others might not; that is, some investors might dislike leverage in the belief it causes tail risks, while others might simply want to avoid explicit leverage on their books.

If we find Gross has no alpha after adjusting for exposures to the market plus the first two factors, or perhaps all three factors, we would be saying that his historical track record is not one most investors should find attractive. If we found, as in FKP, that Gross had alpha but not at the standard levels of statistical significance, we would be saying that his historical track record was that one investor would have preferred to an index, but the difference might plausibly have been luck.

5.4 Gross' skill

Gross definitely had the skills FKP noted above for Buffett: he acquired investors and leverage, he ran his fund efficiently, he stuck with his high-risk principles even when they were going through bad periods, and he communicated so that his investors not only stuck with him, but gave him the funds to build the largest bond fund in the world.

But there is one important difference. FKP explain Buffett's excess returns via factors that were not discovered for many decades after Buffett began investing. But when Gross began in 1987, there was substantial evidence of premia for taking credit risk and negative convexity, and these factors were widely discussed in both practitioner and academic literature. Loading up a portfolio with more credit risk than your index, and running negative convexity versus a positive convexity index, could not be described as skill. Only if Gross had alpha adjusting for those exposures would we call it skill.

The 5s/30s factor is a debatable case. On one hand, many people noted the ability to profit from yield curve rolldown before 1987. On the other hand, the opposite advice, *mutuari brevis longa commodare* (borrow short to lend long), predates the Christian era, and probably the invention of writing. No one had compiled convincing statistical evidence before 1987 that 5s/30s carries a systematic return premium. It was usually presented as an opportunistic strategy to be done based on yield curve shape or state of the business cycle.[2]

Also, while credit and convexity risk are straightforward to obtain (buy BBB bonds, buy mortgages), 5s/30s or similar strategies require models and arguably some degree of skill.

The quantitative difference is that four-factor multiple regressions of Buffett's returns have R^2 values around 0.3, while for Gross the values are around 0.9. The key statistic we use to answer our questions is the t-statistic on the constant term of a multiple regression of manager excess returns on factor excess returns. To a first approximation, errors in the Sharpe ratio of the factor portfolio will cause opposite sign errors in the t-statistic with a proportionality factor of $\sqrt{\frac{R^2}{1-R^2}}$. Buffett's R^2 of 0.3 produces errors in the t-statistics of roughly $\sqrt{\frac{0.3}{1-0.3}} \approx 0.65$. For Gross the corresponding proportionality factor is 3. Given that expected factor returns are very hard to estimate even over long periods, Sharpe ratios are uncertain, and this imposes greater estimation errors on Gross' alpha than Buffett's. Sharpe ratios are sensitive to small changes in factor specification. We discuss below why the high R^2 of Gross' regression leads us to prefer simpler specifications of factors with closer relations to observed market prices than FKP used for Buffett.

Making the same point in words rather than symbols, the volatility of the residual portfolio of Gross minus the regression-estimated factor portfolio is small relative to the volatility of the factor portfolio; so errors in estimating the expected return of the factor portfolio have a magnified effect on the estimated expected return of the residual portfolio. The opposite is true for Buffett.

We try to mitigate this effect by choosing simple factor portfolios whose returns are based directly rather than indirectly on market prices, and also that are similar to what Gross actually did in the hopes that errors in the factor portfolio returns are offset by corresponding errors in Gross' performance. This makes our work less comparable to academic work identifying factor portfolios, but we hope more relevant in evaluating Gross' performance.

6 Results

6.1 Multiple regression

We use ordinary least squares regression to estimate the alpha of PIMCO TR. The results can be found in Table 4.

PIMCO TR is positively correlated to all four factors and all factors are statistically significant at the 5% level. The factors account for 89% of the variance in PIMCO TR monthly excess returns.

Table 4 Regression results over our entire sample period from 1987 to 2014. We show positive loadings at statistically significant levels for all three factors. The model has a high R^2 showing that the model fits the data well. The results in Table 4 indicate that Bill Gross did indeed have alpha in the time he managed the Total Return Fund.

	10-Year Treasury	Credit	MBS	5s/30s	Annualized alpha	R^2
Gross factor exposures						
Coefficient	0.06	0.52	0.21	0.29	0.84%	0.89
Std Error	0.03	0.03	0.06	0.05	0.29%	
t-stats	2.2	20.0	3.8	6.1	2.8	

The regression suggests that Gross had annualized alpha of 0.84% after fees. Since the factors are all shown without fees we should adjust that. If we were looking at Gross' pure investment skill, we would add the fund fees for the share class we used for data. In the more recent years that is 0.42% and it was generally higher in the past, although we do not have precise data. On the other hand, the factors are available in ETF or institutional form for fees on the order of 0.10% to 0.15%, which should be multiplied by the sum of the coefficients, 1.09.

If we were asking whether Gross would have outperformed ETF or institutional factor portfolios exploiting the same return premia, had they existed at the time, our alpha estimate would be above 3. If we were asking whether Gross could have beaten the market without transaction costs, our *t*-statistic could be as high as 4.5. Neither question is of much practical interest except to prove that whatever Gross was doing, it was not throwing darts at a dartboard.

6.2 Comparison to Superstar Investors

Villalon, Brooks and Tsuji (VBT) (2016), in *Superstar Investors* do a similar regression using a more traditional factor approach of a market factor plus zero-investment factors, and find Gross had annualized alpha of 0.32%[3] and a *t*-statistic of 0.94.

Their factors only go back to 1994, so Table 5 compares our factors to theirs on the same time period. Our results are qualitatively similar to our 1987–2014 regression, so the additional time period does not explain the difference in our conclusions about alpha. Our regression matches VBT in R^2, so the standard errors on our alpha estimates are quite close. The difference is that we find a higher estimate of alpha. This implies that our proxy portfolio defined by the regression has a lower Sharpe ratio than the proxy portfolio defined by the VBT regression. We line up the columns to match factors intended to capture similar economic drivers. The big difference in coefficients and standard errors is that since we use three long-only portfolios, market risk is distributed among them, whereas VBT, by construction, concentrate market risk in their first factor. Therefore, only the *t*-statistics are comparable. We find much stronger evidence for credit exposure, somewhat weaker evidence for short volatility, and about the same for low risk. But all *t*-statistics in both regressions are highly significant, with the exception of the 10-year treasury factor in our regression. Over the full 1987–2014 period, Gross had a small but significant exposure to treasuries beyond interest rate exposure

Table 5 Results of our regression in panel A and the results from VBT in panel B. This table compares results over the 1994–2014 time period. The key takeaway is that we find a much higher level of alpha than VBT, a result we attribute to our choice of regression variables, which we feel more closely matches what Bill Gross actually implemented in his portfolio.

	10-Year Treasury	Credit	MBS	5s/30s	Annualized alpha	R^2
Panel A: Our results 1994–2014						
Coefficient	0.01	0.52	0.27	0.41	1.05%	0.88
Std Error	0.03	0.03	0.07	0.06	0.34%	
t-stats	0.21	18.83	3.90	7.35	3.12	
	Market	**Credit**	**Short Volatility**	**Low Risk**	**Annualized alpha**	R^2
Panel B: VBT results 1994–2014						
Coefficient	1.06	0.06	0.04	0.07	0.32%	0.88
Std Error	0.03	0.01	0.01	0.01	0.34%	
t-stats	41.01	6.02	4.53	7.98	0.94	

from our other factors. But from 1994–2014 that was not the case. Another point of interest is that VBT use their long-only exposure of 1.06 to argue that Gross had levered exposure to the market. The sum of our long-only coefficients is only 0.79 suggesting that Gross did not lever market exposure. The difference, of course, is our definitions of the market; VBT use the benchmark, we use a regression-estimated long-only portfolio more similar to Gross' holdings.

Since the two regressions have similar R^2s, the proxy portfolios have similar volatilities, 3.86% for our regression, 3.87% for VBT's. But our proxy portfolio had excess returns of only 3.32% per year, compared to 4.08% for VBT.

This highlights an issue with using multiple regression to evaluate manager alpha. Two sets of researchers using essentially the same approach but making different reasonable choices for factors can come to opposite conclusions on the statistical significance of alpha.

The main problem is that it is hard to estimate expected factor returns even with long data series. There is no academic consensus about the size of the global equity return premium over the last century, so we can hardly put confidence in historical values of risk premia like short volatility fixed-income with 20 years of data.

The solution is to use factors that correspond as closely as possible to what the manager actually did, in the hopes that any deviation in returns over the historical period from long-term theoretical expectation is reflected similarly in the manager's returns and the factor returns.

Fortunately, we can compare our factor selection with VBT because we used the same conceptual factors, just different measurements of them. However, to do an apples-to-apples comparison, we first need to convert each one to a zero-duration, zero-investment version. To do this, we regress each factor return series on the 10-year treasury return. We form a portfolio of the factor, minus the univariate beta on the 10-year treasury, plus or minus enough of the risk-free return to get to zero investment. For a factor that started as zero investment, this means the amount of the risk-free asset is the opposite of the univariate beta on the 10-year treasury–that is, we borrow at the risk-free rate to fund the 10-year treasury beta. For an asset that started as 100% investment, we subtract 1 from that univariate beta–that is, we fund both the factor and its beta on the 10-year treasury.

Our 5s/30s strategy is simple, use futures to go long 5-year treasury exposure and short 30-year, with zero net duration. This precisely matches Gross' description of his strategy, although he accomplished it mainly by overweighting bonds in the 5-year maturity bucket and underweighting bonds in the longest maturity bucket. Although our factor was constructed each month to have zero estimated duration, the resulting factor has a 0.22 beta on the 10-year treasury over this period. The reason is that the factor makes money when the yield curve steepens, and this happens generally when interest rates fall, so the 10-year treasury makes money as well. The duration calculation assumes parallel yield curve shifts, so it does not account for this correlation.

VBT does a more complicated version of the strategy. In their words, "We rank 2-year, 5-year, 10-year, and 20-year US bond futures by their respective durations. The portfolio goes long the futures whose durations are below average, and short the futures with durations above average. Finally, the positions are re-scaled to be duration-neutral." This turns out to have a −0.19 beta on the 10-year treasury, about the same interest rate sensitivity as our factor, but in the opposite direction.

When converted to zero-investment, zero-duration factions, the VBT version has a lower Sharpe ratio, 1.75 versus 2.94. So this does not explain the difference in our conclusions, as it goes in the opposite direction.

For the credit factor, VBT use 5-year US High Yield CDX, while we use the difference between a portfolio with US and international investment grade corporate bonds and duration-matched treasuries. While Gross did use some credit derivatives, nearly all of his credit risk came from holding bonds, and it was nearly all investment grade exposure. He also took the credit risk of international bonds, which differs economically from domestic credit risk (i.e. the spread between AAA and BBB US securities compensates for a different economic risk than the difference between US and European A securities).

Here again the betas of our factors on the 10-year treasury are the same magnitude but opposite sign. Credit has a negative correlation with interest rates, rates fall and defaults rise in recessions, while rates rise and defaults fall in good times. But since our factor is a full investment factor, the impact of interest rates more than offsets the effect of credit.

Putting both on the same zero-duration, zero-investment basis, VBT's version has a much higher Sharpe ratio, 2.14 versus 0.28. Over the 1994–2014 period, betting on high yield credit versus investment grade was eight times as profitable, adjusted for interest rate exposure and volatility, as Gross' credit bets. However, our credit factor has a much higher correlation to Gross' returns, so our factor portfolio holds more of it, reducing the difference from eight times to two times. Stating it a different way, Gross had a levered exposure to investment grade credit rather than an unlevered exposure to high-yield credit.

Overall, credit contributes 0.12% per year to our factor portfolio returns, versus 0.25% for VBT. Of course, positive contribution to factor portfolio returns reduces estimates of Gross' alpha, which in turn reduces estimates of its statistical significance.

To capture the short volatility factor, we use the return on a portfolio of mortgages. VBT use 1-month, 30-delta strangles on 10-year treasury futures. In fact, Gross got most of his short volatility or negative convexity from mortgages. He did use options on treasury futures, but their effect was smaller and they do not seem to correspond to strangles (although we admit we cannot analyze the positions closely). Mortgage convexity comes from prepayment risk, which is correlated with interest rate volatility but not identical to it.

This is the biggest reason for our differences from VBT. Mortgages adjusted for treasury exposure actually had a negative Sharpe ratio over the period, −0.68, while strangles had the highest positive Sharpe ratio of any factor either of us used, 4.53. Once again our factor has a much higher correlation but the overall effect is our factor portfolio lost 0.06% per year due to being short volatility by owning mortgages, while VBT's made 0.43% per year by buying strangles.

Here again we feel that our factor captures what Gross actually did better. Adjusting his alpha for things he could have done but did not, like writing credit protection on high-yield bonds or writing strangles on the 10-year treasury, seems contrary to the spirit of the questions we are asking. VBT asked somewhat different questions, so we do not critique their choices.

We are now in a position to approximately reconcile the two sets of results. Gross' single-factor alpha versus benchmark was an annualized 1.32% over the 1994–2014 period, with a t-statistic of 3.42 (the 1.33% mentioned at the beginning of the paper is for the full 1987–2014 period). The univariate adjustments for our three factor portfolios were −0.46% for yield curve, −0.12% for credit and 0.06% for short volatility. $1.32\% - 0.46\% - 0.12\% + 0.06\% = 0.80\%$, close to the 0.84% alpha we find for Gross. The difference between 0.80% and 0.84% arises because summing univariate contributions does not exactly equal multivariate contributions, and also because the estimates use different time periods.

For VBT the corresponding values are −0.29% for yield curve, −0.25% for credit and −0.43% for short volatility. $1.32\% - 0.29\% - 0.25\% - 0.43\% = 0.36\%$, close to the 0.32% alpha VBT finds. In this case the difference is entirely due to interaction effects among the factors as both numbers are derived using the same time periods. The ratio of our alpha numbers translates closely to the ratio of our t-statistics, since the standard errors of our regressions are nearly identical.

6.3 Timing alpha

The multiple regression captures only average exposures to market factors over the entire 1987–2014 period. Another source of alpha is timing, that is, varying factor exposures.

To investigate this for each of our four factors, we divide our sample into months the factor had a positive return, and months the factor had a negative return. We then regress PIMCO TR excess returns on factor returns over the entire sample and in both subsets. In addition to the four factors individually, we show results for the linear combination from the multiple regression that forms the proxy portfolios of factors that best matches Gross' returns.

The results are shown in Table 6.[4]

Interestingly, Gross exhibited the same pattern for each factor and the proxy portfolio. His timing ability was negative, with the marginal exception of Credit. He had higher factor exposure when

Table 6 PIMCO TR Factor Exposures group by whether the month was a positive return (up month) or negative return (down month) for the factor. The sample period is for the full history from 1987 to 2014. The R^2 in up months and down months are similar across our factors showing that the model is not simply capturing a correlated move in an up or a down market. The 5s/30s factor displays positive alpha in both up and down markets, which is not surprising as it is immunized from parallel moves in the level of interest rates.

	10-Y Treasury	Credit	MBS	5s/30s	Proxy portfolio
Entire sample					
Factor exposure beta	0.57	0.74	1.12	1.08	1.00
Std Error	0.02	0.02	0.04	0.04	0.02
Annualized alpha	2.00%	1.33%	0.57%	2.77%	0.84%
R^2	0.69	0.82	0.71	0.38	0.89
Factor down months					
Factor exposure beta	0.59	0.66	1.12	0.98	0.98
Std Error	0.06	0.05	0.10	0.19	0.05
Annualized alpha	1.74%	−0.65%	−0.62%	0.73%	0.25%
R^2	0.44	0.62	0.57	0.15	0.76
Factor up months					
Factor exposure beta	0.47	0.70	0.94	0.77	0.94
Std Error	0.04	0.04	0.07	0.14	0.04
Annualized alpha	4.18%	2.51%	2.85%	6.11%	1.85%
R^2	0.44	0.64	0.41	0.15	0.73

the factor went down than when it went up. That could be from bad timing, increasing exposure before down months and decreasing it before up months, or it could result from negative convexity in that factor.

On the other hand, Gross generated lots of positive alpha in up months, and much smaller or negative alpha in down months. Whatever Gross' strategies were for alpha generation after adjusting for our four factors, they worked when factors were up, but not when factors were down.

If you imagine a graph of Gross' returns versus any of the factor returns, the points to the right of the y-axis, the months with positive factor returns, have a regression line with shallower slope and higher intercept than the points to the left of the y-axis.

We investigate the timing more formally with a Henrikkson–Merton regression.[5] For each factor plus the proxy portfolio, we regress Gross' excess returns on both the factor, and the factor returns with negative months set to zero. The beta coefficient on the latter term is a measure of timing skill, a positive value means Gross profited from timing (or from positive convexity, we cannot tell the difference with these data), a negative value means Gross' returns suffered from poor timing or negative convexity.

We convert this beta into a return equivalent by multiplying it by the annual cost of 1 month

Table 7 Results of running a Henrikkson–Merton Regression on the factors over the full sample set from 1987 to 2014. The Henrikkson–Merton Regression provides a proxy for the excess return that is not explained by a mix of options and the market portfolio. We present the factor exposure beta, followed by the factor option beta. The third subtable shows the option value and indicates that Gross mostly lost money from his market timing decisions as evidenced by the negative alpha on the proxy portfolio and three of the four factors.

	10-Y Treasury	Credit	MBS	5s/30s	Proxy portfolio
Factor exposure beta	0.64	1.29	0.73	1.26	1.04
Std Error	0.04	0.09	0.04	0.16	0.04
t-stat	14.66	15.11	19.51	7.92	25.12
Factor option beta	−0.14	−0.28	0.03	−0.31	−0.07
Std Error	0.07	0.13	0.06	0.24	0.06
t-stat	−2.02	−2.23	0.82	−1.31	−1.06
Annualized option value	−1.17%	−1.23%	0.22%	−1.03%	−0.37%
Annualized alpha	3.14%	1.74%	1.12%	3.73%	1.21%
Sum	1.97%	0.51%	1.34%	2.70%	0.83%
R^2	0.69	0.71	0.71	0.38	0.89

at-the-money options on the factor, using Black–Scholes value assuming a constant volatility equal to the actual volatility over the period. The result is an estimate of the option value generated or lost (in Gross' case, mostly lost) through timing decisions and convexity.

The Henrikkson–Merton results are only similar to the previous analysis at first glance. It is true that factor option betas are negative for four of the five factors. But they are significant at the 5% level only for 10-year treasury and Credit, whereas the biggest differences between up-month and down-month betas were for MBS and 5s/30s.

Moreover, while the annualized option values are large for three factors, meaning that it would have cost a lot of money to buy options to offset Gross' negative timing decisions or negative convexity, on the proxy portfolio the annualized option value is relatively small. It is of course a mathematical fact that an option on a portfolio has a value less than or equal to options on its constituents, but in Gross' case the effect is surprisingly large. So Gross' interest rate exposure went up when yields were increasing, and his credit exposure went up when credit spreads were increasing, but total exposure as best we can measure it went up a significantly smaller amount in bad months for the combined factors.

The mathematical difference between the up/down month regressions and Henrikkson–Merton is the former allows different intercepts for positive and negative factor return months, while the latter forces them to the same intercept. So the up/down regression finds more negative timing or convexity, but also more alpha in positive months. Henrikkson–Merton combines those effects to find much weaker evidence for negative timing or convexity.

One explanation that has often been offered for Gross' alpha is that the size of his fund and the quality of the PIMCO operation allowed him to get leverage and access to new issues on preferential terms. If so, we might speculate that these advantages matter more in factor up markets. If bond prices are going up, or long credit, short convexity or short volatility are paying off, there could be more competition for these exposures, and the fund with the most market power would have an advantage. When the market or factors are not doing well, there could be less competition to acquire them, both because managers are losing assets and because managers are cutting exposures to the underperforming factors. Preferential access is valuable when customers are clamoring for exposures, not when customers are scarce.

7 Conclusion

We find significant positive alpha for Gross, but we cannot disprove the negative result offered by VBT. We (naturally) prefer our choices of factors, but not to the extent of asserting our results are correct. It is true that we use a longer data series, so we could find significance where VBT do not due to less parameter uncertainty, but when we compare for identical sample periods, we still find significant alpha and VBT do not. Until better research comes along, we conclude that Gross appears to have demonstrated alpha over 1987–2014 period, but that its statistical significance depends on the choice of factors. However, if you adjust factors for fees or add fees back to Gross' returns, our results become more strongly statistically significant and VBT would near or approach the conventional 5% threshold.

We want to make clear that even if Gross's performance can be explained in some, or even in large part by these factors, that this should not undermine his accomplishments as an investor. Gross' ability to stick to these strategies, identify low cost and consistent sources of leverage and effectively communicate his strategy to a dedicated investor base is impressive. We hope this research adds to the growing discussion on factors in fixed-income markets and the literature on measuring investor alpha.

Acknowledgments

We'd like to thank Sanjiv Das and Sameer Kackar for their helpful comments.

Notes

[1] The TRACE database was introduced in 2002 and could have been used for a portion of our sample, however for simplicity we choose not to use it. Also, while it was a considerable advance in transparency in bond market transparency, it falls well short of consolidated equity trade data in quality.

[2] In other words, as a steepener bet, not a strategy to earn rolldown.

[3] The published paper rounds to 0.3%, but VBT were kind enough to supply us with their data so we could replicate their findings.

[4] It may seem odd that the full sample exposures are not between the down and up month exposures, except for the 10-year treasury. If you imagine the graph of fund exposure versus factor exposure, you can think of the full sample exposure as mainly determined by the slope between the center of mass of points to the left of the y-axis and the center of mass of the points to the right of the x-axis. This has no necessary relation to the slope of points within either region. Only if there were a strong causal relation between x and y would we expect the full sample exposure to be in between the two subsample exposures.

[5] Henrikkson and Merton (1981).

References

AQR Capital Management (2017). "The Illusion of Active Fixed-Income Diversification," *Alternative Thinking*.

AQR Capital Management (2016). "Style Investing in Fixed Income," *Alternative Thinking*.

Asness, Clifford S. (1995). "The Power of Past Stock Returns to Explain Future Stock Returns," *Goldman Sachs Asset Management*.

Asness, Clifford S., Andrea Frazzini and Lasse Pedersen (2018). "Quality Minus Junk" *Review of Accounting Studies*.

Asness, Clifford S., Andrea Frazzini and Lasse Pedersen (2012). "Leverage Aversion and Risk Parity," *Financial Analysts Journal* **68**(1).

Brooks, Jordan. Tobias J Moskowitz (2017). "Yield Curve Premia," *Working Paper available on SSRN*.

Burham, Benson, J. (2016). "Betting Against Beta with Bonds," *Financial Analysts Journal* **72**(6).

Carhart, M. M. (1997). "On Persistence in Mutual Fund Performance," *The Journal of Finance* **52**, 57–82.

Cochrane, John and Monika Piazzesi (2005). "Bond Risk Premia," *The American Economic Review* **95**(1).

Fama, Eugene F. and Kenneth R. French (1992). "The Cross-Section of Expected Stock Returns," *The Journal of Finance* **2**, 427–465.

Ferson, Wayne E. and Campbell Harvey (1991). "The Variation of Economic Risk Premiums," *Journal of Political Economy*.

Frazzini, Andrea, David Kabiller and Lasse Pedersen (2018). "Buffett's Alpha" *Financial Analysts Journal*.

Frazzini, Andrea, Lasse Pedersen (2014). "Betting Against Beta," *Journal of Financial Economics* **111**(1).

Gross, William H. (2005). "Consistent Alpha Generation through Structure," *Financial Analysts Journal* **61**(5).

Henrikkson, R. and R. Merton (1981). "On Market Timing and Investment Performance: Statistical Procedures for Evaluating Forecasting Skills," *Journal of Business* **54**(5), 513–533.

Jurek, Jakub and Erik Stafford. "The Cost of Capital for Alternative Investments," *The Journal of Finance*, October, 2015.

Lo, Andrew and Peter Lee (2014). "Hedge Fund Beta Replication: A Five Year Retrospective," *Journal of Investment Management* **12**, 5–18.

Ilmanen, Antti (1995). "Time-Varying Returns in International Bond Markets." *The Journal of Finance*.

Litterman, Robert and Jose Scheinkman (1991). "Common Factors Affecting Bond Returns," *The Journal of Fixed Income* 54–61.

Shiller, Robert and John Campbell (1991). "Yield Spreads and Interest Rate Movements: A Bird's Eye View," *Review of Economic Studies*.

Villalon, Daniel, Jordan Brooks and Severin Tsuji (2016). "Superstar Investors," *The Journal of Investing*.

Keywords: