Takahē Capital Insights

TAKAHĒ
www.takahe.capital
contact@takahe.capital

October 2024

Worth a Hill of Means

By Moritz Seibert

Worth a Hill of Means

Takahē Capital maintains a large database of quantitative strategies to support and improve its research process. While some of these strategies were developed in-house, most were sourced from books, academic papers, and the internet. One such example is the trading system featured in this Insight note: a long-biased, short-term mean reversion system trading the S&P 500 e-mini futures contract.

We started tracking this system in August 2014 after discovering it in a research note published by Quest Partners LLC. It is now approaching its 10-year live anniversary in our database.

This Insight note reviews its performance and highlights some interesting results and conclusions.

What to expect from the strategies in our database

Most of the strategies included in our database[1] are either ill-designed or have experienced substantial performance decay over time. Once the rules of a trading system are made public, its excess returns tend to decline in response to broader adoption.[2]

While many academic papers present systems that lack realistic trading assumptions and therefore look too good on paper, our default assumption is that the strategies shared by active traders are mediocre at best and useless at worst. That's because professional traders are motivated to protect their IP and unlikely to compromise their competitive advantage by revealing the details of their strategies.

Despite these limitations, we are convinced it's valuable to follow a broad spectrum of these external strategies and shadow-run them in parallel to our own. If nothing else, our database creates a benchmark and a frame of reference from which we can learn.

Mean reversion

Opposite to trend following strategies, mean reversion systems adopt a convergent approach by exploiting price oscillation around the mean of the return distribution rather than price excursions into the tails.

Generically, these systems are designed to buy dips and/or sell rallies – a method that tends to produce a large number of small winning trades and fewer, but larger, losing trades. Kahneman and Tversky have shown that this positive win vs. loss relationship is emotionally more palatable for most investors, whereas trend following systems can induce emotional pain because losing trades tend to outnumber the winning trades, resulting in drawdowns and a postponed sense of success. However, mean reversion systems are prone to suffer to a greater extent from price discontinuities, also known as gaps, producing a negatively skewed trade and return distribution.

And they can very easily be over-optimized.

BTFD

The Quest note focuses on the illustration of irregular returns in US equity and bond markets. Quoting from their article: "By illustrating the availability of skill-less and unstable mean reversion based sources of alpha, we hope investors will exercise sincere self-analysis of

[1] At the top-level, our database categorizes strategies as follows: trend, countertrend, carry, mean reversion, spreads, patterns, and setups.

[2] Zhou, Lin (2017): The alpha life cycle of a quantitative strategy. Falck, Rej, Thesmar (2022): When do systematic strategies decay. McLean, Pontiff (2015): Does academic research destroy stock return predictability.

their methods of investment selection and also of the fees that they are willing to pay their managers." And: "We do not trade such models, nor do we recommend for anyone to trade them as we believe it is only a question of time before substantial losses are experienced in these or similar strategies."

We couldn't agree more.

The system presented in their article is a long-sided mean reversion system whose trading parameters were intentionally optimized for the S&P 500 e-mini and US 30Y treasury bond futures contracts. In a nutshell, the system buys the market on the third down day if it has fallen more than 5%[3] of its 50-day average true range (ATR)[4] on three consecutive trading days.

The system has three exit rules, closing the long position on whichever event happens first:

- A *time exit* which sells market at settlement on the 4th business day since entry.
- A *stop loss exit* which sells on a stop equal to the entry level minus 2 ATRs.
- A *profit target exit* which sells on a limit equal to the entry price plus 0.5 ATRs.

Ten years ago, when we read the Quest paper, we thought that this simple and over-optimized mean reversion system must very soon stop working and deteriorate. Well, we were wrong – or maybe not right yet.

Figure 1 shows the in-sample performance of this system between 1 January 2004 and 31 July 2014, trading S&P e-mini futures.[5] The green line depicts the performance with zero slippage and zero commissions; the blue line includes 2 ticks of slippage ($25) and $5 commission per executed contract. The tick size for the S&P 500 e-mini futures contract is 0.25 and the value of 1 tick equals $12.5.

The system always trades 1 contract, and our account has a starting equity of $100,000. Margin interest is not included.[6]

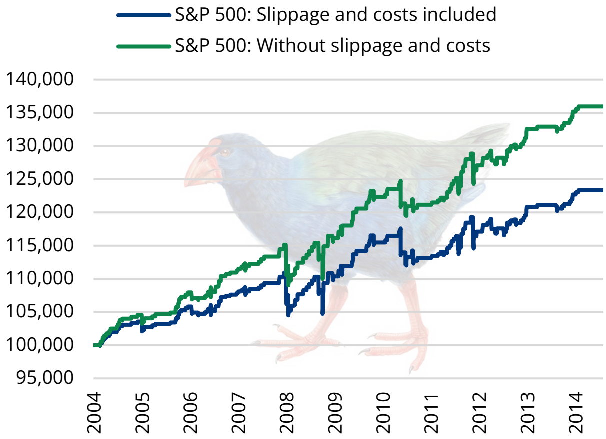

Figure 1: In-sample performance trading S&P 500 e-mini with and without trading costs.

Figure 2, shown on the following page, illustrates the in-sample performance for the same system and time period, this time trading US 30Y treasury bond futures. Commissions are $5 and slippage is 1 tick ($31.25) per executed contract.

Again, the account starts with $100,000, the trade size is always 1 contract, and margin interest is excluded.

---

[3] The system works essentially identical if this 5% condition is removed. For convenience, the version which excludes this condition is the one we analyze in this Insight note.

[4] The true range (TR) for day t equals the maximum of today's high and yesterday's close less the minimum of today's low and yesterday's close, thereby accounting for overnight price discontinuities:

$TR_t = \max(High_t\,, Close_{t-1}) - \min(Low_t\,, Close_{t-1})$

The average true range (ATR) is the n-day average of the true range. For the buy-the-dip system, n=50 days and the average is calculated arithmetically.

[5] Data is sourced from Commodity Systems Inc. (CSI) for CME Globex S&P 500 e-mini futures. The time series is backadjusted for price differences at the time of the rollover, and contracts are switched at settlement two days prior to expiration.

[6] We get very similar results when running this system on the S&P 500 SPDR (SPY) ETF.

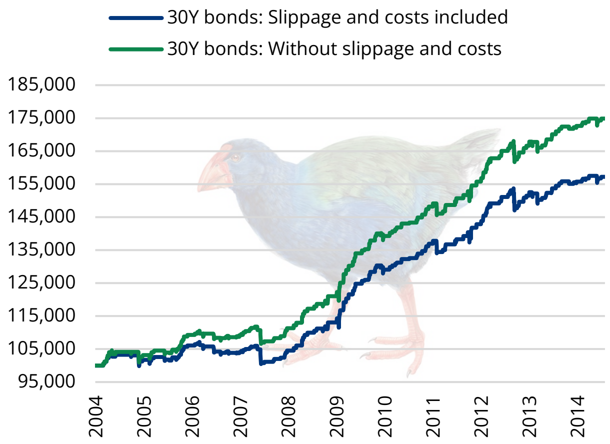

Figure 2: In-sample performance trading US 30Y treasury bond futures with and without trading costs.

Next, in Figure 3, we attach the out-of-sample performance between 1 August 2014 and 31 July 2024 to the in-sample results, providing a visual comparison of the backtest and live results.

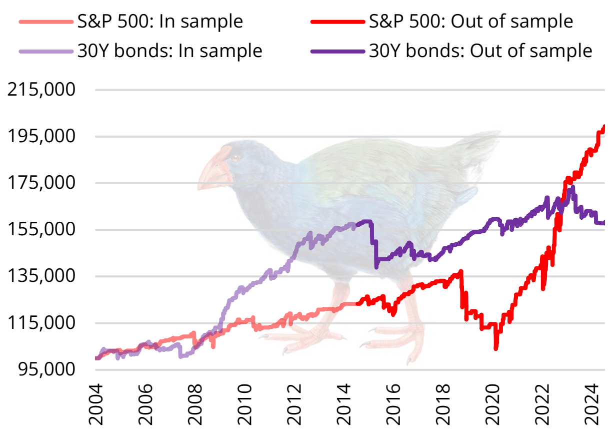

Figure 3: Out-of-sample performance added to the in-sample results. Costs are included.

The system trading US 30Y bond futures, shown in purple, had a drawdown at the start of the live period around 2015, followed by a recovery and a new decline which continues to the present day, most likely caused by an environment of rising interest rates and falling bond prices. In contrast, while the S&P 500 e-mini system, shown in red, experienced a drawdown during the COVID period, its subsequent returns accelerated, propelling the equity curve to a new all-time high.

There are several potential explanations for the recent improved performance of the S&P 500 e-mini system, such as the persistent and relentless bid on equity index ETFs from passive investors, central bank policies which provide liquidity and support asset prices, and so on. However, pinpointing an exact cause is challenging if not impossible. Instead, we prefer to let the data speak for itself.

When trading both markets in parallel to create a small BTFD portfolio, we get the performance results shown in Figure 4.

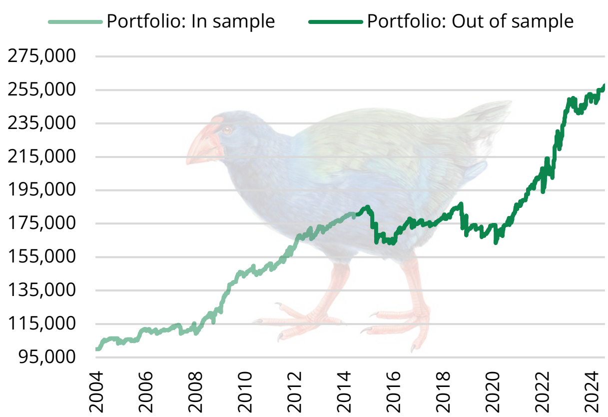

Figure 4: Portfolio performance. Costs are included.

The portfolio looks attractive, supported by the low (essentially zero) correlation between the individual return streams. However, before fully committing to this system, we believe it's essential to consider some relevant questions:

- How does the same system work on the short side?
- Does the system generalize well across other markets?
- How robust is the system and how sensitive are the results to changing parameter values?
- In the case of the S&P 500 e-mini system, would similar results be achieved if we naively shorted VIX futures or S&P 500 puts?

Table 1 shows some relevant trade statistics, including the in-sample, out-of-sample, and full-sample periods. Costs are included.

Statistic[7]	**S&P 500 e-mini**			**US 30Y bonds**		
	In Sample	**Out of Sample**	**Full Sample**	**In Sample**	**Out of Sample**	**Full Sample**
CAGR	1.49%	5.63%	3.28%	2.36%	-0.14%	1.15%
Sharpe Ratio	1.01	1.11	0.95	0.97	N/A	0.4%
Worst Drawdown	-3.89%	-9.31%	-9.31%	-5.83%	-10.31%	-10.31%
Number of Trades	125	135	260	142	89	231
Total Winning Trades	98	113	211	102	55	157
Total Losing Trades	27	22	49	40	34	74
% Winning Trades	78%	84%	81%	72%	62%	68%
Average Winner	0.38%	0.8%	0.54%	0.6%	0.75%	0.62%
Average Loser	-0.69%	-1.59%	-0.99%	-0.78%	-1.23%	-0.89%

Table 1: Trade statistics for the buy-the-dip system.

Trading the short side

Rather than buying dips, this section presents an inverted system designed to sell rallies. All trading parameters are unchanged but adapted to work on the short side. However, to prevent a negative balance, the initial account equity is increased to $1 million. Figure 5 shows the performance for the S&P 500 e-mini futures.

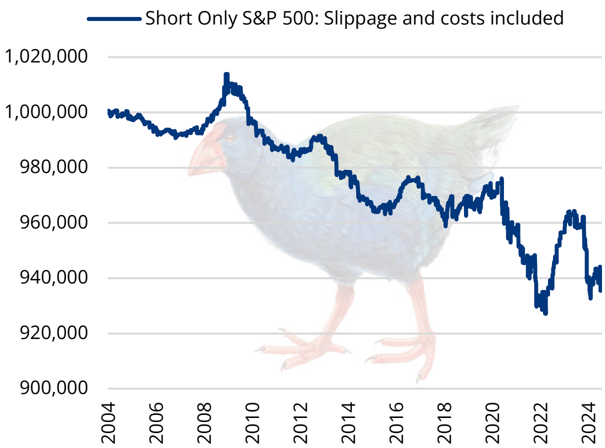

Figure 5: Short-only S&P 500 system.

Figure 6 depicts the results for the short-only version trading the US 30Y treasury bond futures.

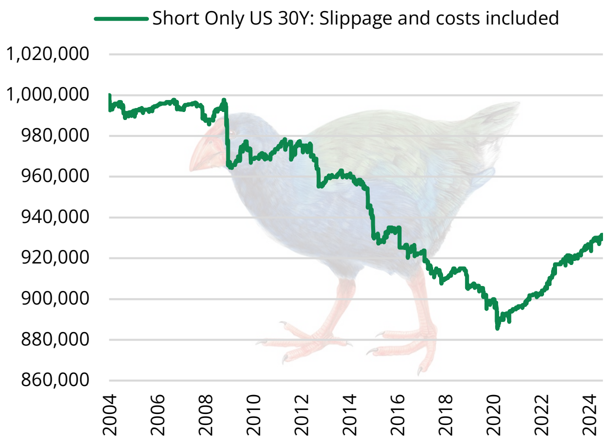

Figure 6: Short-only US 30Y system.

Though the short-sided S&P 500 system would have generated gains during the Global Financial Crisis (GFC) and at the onset of the war in Ukraine, its overall performance has been disappointing. In contrast, the returns of the US 30Y system appear to depend on the general

[7] Note that the magnitudes of certain statistics, e.g., Average Winner and Worst Drawdown, are larger out-of-sample than in-sample, which is due to larger notional contract values in the out-of-sample period. The average index level of the S&P 500 was greater out-of-sample than in-sample, leading to larger gains and losses when assuming a constant 1-lot trade size. The main takeaway is that the out-of-sample system results for the S&P 500 e-mini have increased whereas for the US 30Y treasury bonds they have decreased.

trend in bond prices. Around 2020, as bond prices started declining in response to rising interest rates, selling short-term rallies would have made money. A motivated researcher might consider adding a directional filter, such as a moving average, to conditions this to buy dips only when bond prices are above this moving average, and vice versa.

Testing other markets

Since buying dips in the S&P 500 index seems to make money, shouldn't the same logic also work well in other index markets?

Figure 7 illustrates the performance of the long-sided system trading several other equity index futures markets. The system always trades 1 contract. Slippage and costs are included.

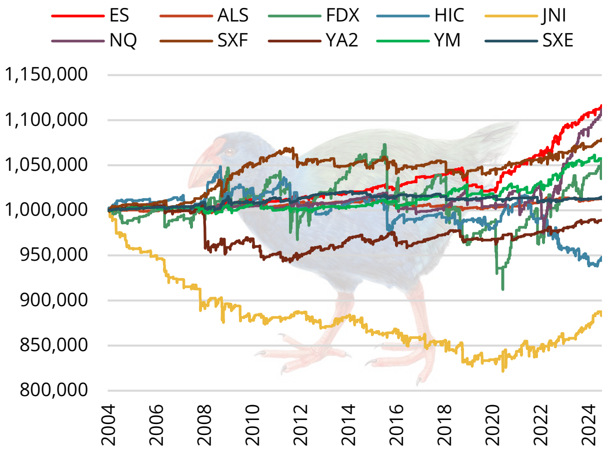

Figure 7: Long-only system trading other equity index futures markets.[8]

While the system would have worked relatively well on North American futures markets such as the S&P 500, the Nasdaq, the Dow 30, or the S&P Canada 60 index, it would have lost money trading Asian markets and performed choppy in most European markets.

The hypothetical performance of an equity index portfolio trading the 10 single equity index markets is shown in Figure 8 below. Again, the position size is 1 contract, and no adjustments are made for volatility or different notional contract values.[9]

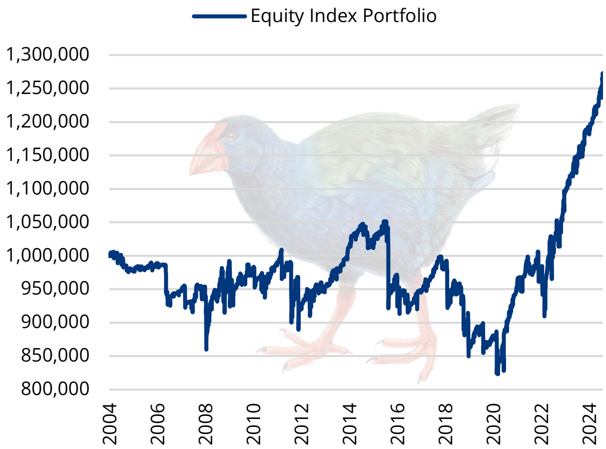

Figure 8: Performance of an equity index portfolio.

In contrast to the period prior to 2020, the chart in Figure 8 indicates that buying equity dips post COVID would have worked quite well.

But can the same effect also be observed in markets that belong to other asset classes, e.g., in crude oil, corn, or the Japanese Yen?

Figure 9 exhibits the performance for various currencies, commodities, and bonds. Slippage and costs are included, and the positions size is again 1 contract. All trading parameters are kept unchanged, and the system only buys dips.

Gold (GC2) and copper (HG2) are the only markets that would have generated gains; all other markets would have produced losses.

[8] ES: S&P 500. ALS: FTSE/JSE Top 40. FDX: DAX. HIC: Hang Seng. JNI: Nikkei 225. NQ: Nasdaq 100. SXF: S&P Canada 60. YA2: SPI 200. YM: Dow Jones 30. SXE: Eurostoxx 50. All data is sourced from Commodity Systems Inc. (CSI). The time series are backadjusted for price differences at the time of the rollover, and contracts are switched at settlement two days prior to expiration.

[9] Adjusting position sizes as a function of contract values and ATR or volatility estimates is something most traders would do in practice. However, the aim of this section isn't the creation of a perfect backtest, but to show that the raw system doesn't generalize well across markets.

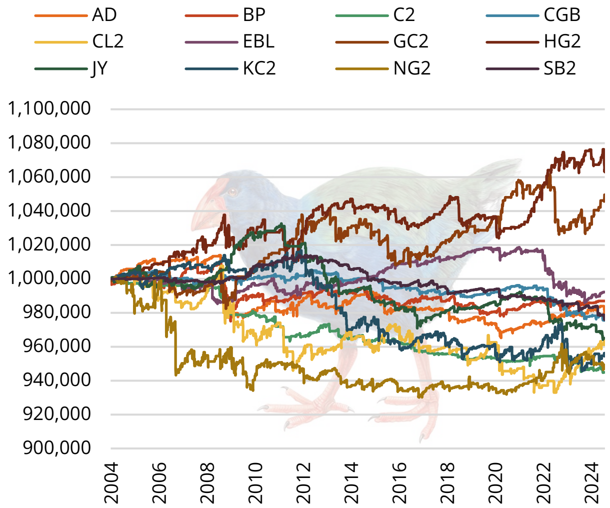

Figure 9: Long-only system trading futures markets from sectors other than equities.[10]

It is a sign of instability when systems don't generalize well across different markets. While it may be tempting to come up with different parameter settings for each market and side, such procedure will only deteriorate robustness and result in a reduced and inconsistent sample size from which valid statistical inferences can no longer be drawn.

Parameter robustness

Another indication of over-optimization and system instability is when its performance deteriorates substantially in response to minor parameter changes.

The surface chart in Figure 10 shows the system's Sharpe ratio for different profit target and stop loss combinations. The red-highlighted peak area corresponds to a Sharpe ratio of about 0.9 for the long-sided S&P 500 system with a 2 ATR stop loss and a 0.5 ATR profit target.

As expected, since the system was intentionally optimized to work for the S&P 500 index, different stop loss and profit target combinations result in a significant performance deterioration.

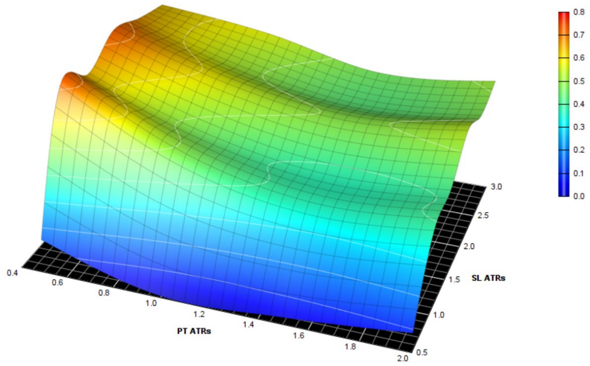

Figure 10: Sharpe ratio of the S&P 500 system for different combinations of stop loss and profit target ATR offsets.

Since this system has a maximum hold period of 3 bars (days) per trade, and because it is designed to exploit a short-term price sequence, additional robustness tests such as variance or noise testing are less relevant. For systems with unconstrained hold periods, such as long-term trend following strategies, these tests can offer valuable insights and an additional analysis dimension to the research process. By altering the sequence of historical returns without changing the descriptive statistics of the dataset, adding small amounts of noise to daily prices, or testing the system's sensitivity to delayed entries and exits, researchers can better understand a system's behavior and risks, and contextualize the odds of larger future drawdowns relative to what the raw historical data provided.

Conclusions

Most strategies in our database are over-optimized in one way or another. Being able to separate the good/resilient from the bad/fragile ones is a central aspect of our work at Takahē Capital. In tennis, matches can be won by avoiding unforced errors, and in trading a lot of money can be saved by staying clear of instable systems.

Heavily over-optimized systems, like the mean reversion system presented in this Insight note,

[10] AD: Australian Dollar. BP: British Pound. C2: Corn. CGB: Canadian 10Y bonds. CL2: WTI crude oil. EBL: German Bunds. GC2: Gold. HG2: Copper. JY: Japanese Yen. KC2: NY coffee. NG2: Natural gas. SB2: NY sugar. All data is sourced from Commodity Systems Inc. (CSI). The time series are backadjusted for price differences.

usually decay quickly out of sample and rarely return to profitability. However, this S&P 500 system is an exception. Not only did it avoid performance deterioration, but it also managed to generate higher returns out of sample.

From a statistical viewpoint, it is expected to encounter outliers in any large enough dataset, including our database. Yet, it appears that buying dips in the S&P 500 has more edge than a random selection from our database would indicate – at least for now.

When analyzing strategies from our database, it is valuable to benchmark them against existing risk premia strategies, factor indices, or smart beta ETFs, as this can sometimes give us a better understanding of the return drivers that are at work.

In this case, a comparison with the CBOE S&P 500 PutWrite Index (PUT) or a system selling VIX futures seems appropriate.

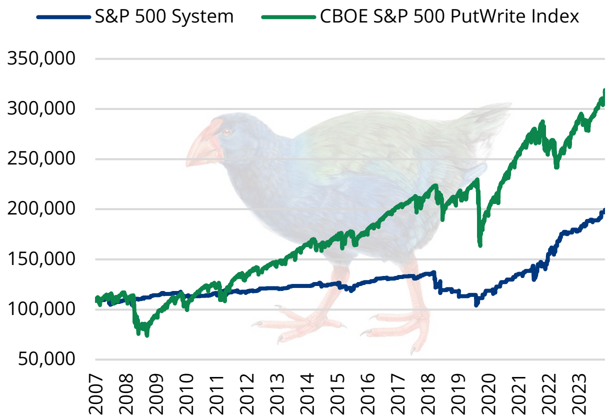

Figure 11: S&P 500 system vs. CBOE S&P 500 PutWrite Index, unadjusted for volatility. Source: CBOE website, www.cboe.com.

Although the returns of both strategies are far from identical, it is no surprise that they tend to win and lose around the same time. Notwithstanding the degree of over-optimization, the key question for investors is whether they want these negative skew strategies included in their portfolio in the first place, given their tendency to exacerbate losses during times of market stress.

About Takahē Capital

Takahē Capital is a private, quantitative investment manager based in Germany. We trade a diverse and evolving set of systematic trading strategies, aiming to generate uncorrelated performance well ahead of the return on cash. Our strategies are resilient in design and extend over multiple trading frequencies and markets. For more information, please visit www.takahe.capital.

DISCLAIMER AND IMPORTANT LEGAL INFORMATION

This document is for informational and marketing purposes only. It is addressed to professional investors and not for public distribution. The information contained in this document must not be copied, reproduced, republished, or posted – in whole or in part – without the prior written consent of the author. Any information and opinions contained herein have been compiled or arrived at in good faith based upon information obtained from sources believed to be reliable. However, such information has not been independently verified and no guarantee, representation or warranty, express or implied, is made as to its accuracy, completeness or correctness. All such information and opinions are subject to change without notice. This document does not constitute an offer to sell or a solicitation of an offer to buy an interest in any fund, account, financial instrument, or investment vehicle. Any such offer, sale or solicitation may be made only pursuant to a confidential private placement memorandum, limited partnership agreement, investment management, and legal subscription documents. All Takahē Capital investment strategies involve a high degree of risk and leverage and there can be no assurance that investors will receive a positive return on their capital. In addition, the performance of all Takahē Capital strategies is volatile and will likely vary substantially on a monthly, quarterly, or annual basis. Any performance information included in this document is not a measure of actual returns, not based on audited financial statements, and dated. The performance results may have decreased since the issuance of this document. ***Past performance is not indicative of future results.***