A risk-based explanation of cryptocurrency returns

Mykola Babiak[†] Daniele Bianchi[§]

Abstract

We investigate the dynamics of returns in cryptocurrency markets through the lens of a small-scale latent factor model with time-varying factor loadings instrumented by individual cryptocurrency characteristics. We have three main empirical findings. First, our dynamic factor model excels in providing a risk-based explanation of daily realised and expected returns across cryptocurrencies, improving over both static latent factor models and pre-specified portfolios sorted on observable characteristics. Second, we show that expected returns are primarily driven by liquidity, volatility, and past performance. Third, our model provides evidence of an increasing, although limited, spill-over of fundamental risk factors between equity and cryptocurrency markets.

Keywords: Instrumented PCA, Cryptocurrency markets, Asset pricing, Mean-Variance efficiency

JEL codes: G11, G12, G17, C23

[†]Department of Accounting and Finance, Lancaster University Management School, UK. E-mail: m.babiak@lancaster.ac.uk Web: sites.google.com/site/mykolababiak.

[§]School of Economics and Finance, Queen Mary University of London, UK. E-mail: d.bianchi@qmul.ac.uk Web: whitesphd.com.

1 Introduction

It is a fundamental tenet of asset pricing that investors should be compensated for their exposure to sources of systematic risk. This principle is intimately related to the quest for return predictability and market efficiency. If the return premium associated with a given asset arises because it is fundamentally riskier, then we might expect these return premiums to persist in the future. If, however, the returns premiums do not reflect a compensation for risk, then we might expect the excess returns to be vanish over time as investors become more aware. This practical view applies to any risky asset and is often based on the assumption that an accurate identification of common factors in the cross section of returns can help to provide a risk-based explanation of the performance across assets (see, e.g., Giglio and Xiu, 2021).

With the rising price and public awareness of Bitcoin, investors have been drawn to cryptocurrency markets by the promise of significant returns, compared to the paltry yields often on offer from cash, bonds, and other traditional asset classes. The hyperbolic growth in notional value – with a total capitalisation of around \$1 trillion at the time of writing – has led investors and academics to more carefully examine the interplay between risks and returns in what is still a relatively unknown market.[1] Put it differently, the extent to which cryptocurrency returns are consistent with the exposure to sources of risk, or if they represent primarily a behavioural phenomena, is yet to be fully understood. This is partly due to the still relatively unknown nature of the risks that market participants face when investing in this space. Following the Fama and French (1993) blueprint for equities, Liu et al. (2022) and Cong et al. (2021b) propose a series of long-short portfolios based on different asset characteristics – for instance, market capitalization, network growth, or past performances – to rationalise part of the variation in cryptocurrency returns.

While this certainly simplifies an empirical analysis, the assumption that risks can be unequivocally mapped by observable factors requires a previous complete understanding of the cross section of average cryptocurrency returns. However, this is likely a partial understanding at best, and the ubiquitous presumption that risk factors can actually be observed with negligible measurement error may not necessarily hold in practice (see, e.g., Giglio and Xiu, 2021). Furthermore, standard pricing models based on either latent or observable common factors typically assume that loadings are con-

[1] For comparison, as of September 2022, the total market capitalization of the Italian, Spanish, French and German equity markets was approximately \$0.7, \$1, \$2.4, and \$1.9 trillion, respectively.

stant over time. As a result, they may be ill-suited for estimating a stochastic discount factor which may not necessarily be constant over time, particularly when the asset class under investigation is plagued by extreme volatility and structural instabilities due to widespread market fragmentation (see, e.g., Makarov and Schoar, 2020, 2021).

In this paper, we build upon Kelly et al. (2019, 2022) and implement an instrumented principal components analysis (IPCA); a conditional latent factor model in which the factor loadings are driven by a set of observable individual asset characteristics. These individual characteristics capture key risk features, including trading frictions, liquidity, volatility, past performances, and growth/adoption as proxied by on-chain network activity.

Our modeling framework has a few features that differ from frameworks used in the prior literature on cryptocurrency markets. To begin, long-short portfolios in the cryptocurrency space do not necessarily represent actual investment opportunities, because they hardly incorporate transaction costs and trading restrictions (see, e.g., Makarov and Schoar, 2020). Discrepancies between the construction of factor portfolios and their actual implementation could bias inferences about a beta/expected return model (see, e.g., Huij and Verbeek, 2009). By acknowledging that the common factors are unobserved, our approach searches for the most apt factors and avoids theoretical inconsistencies from fixing factors *a priori* and treating them as though they are perfectly observed.

In addition, IPCA implies that the factor betas are time varying and depend on a potentially large set of individual asset characteristics. On the one hand, this allows a researcher to discipline the estimates of the stochastic discount factor in a way that is coherent with theoretical underpinnings; that is, individual asset characteristics should provide reliable information to understand the dynamics of expected returns (see Daniel and Titman, 1998). On the other hand, the instrumented betas can help to consistently recover the unknown factor structure of the returns while at the same time can accommodate dramatic fluctuations in the pricing kernel (see, e.g., Cochrane, 2011; Kelly et al., 2019).

As a by-product of the estimation framework, the IPCA produces a set of managed portfolios, one for each characteristic, that can be used to test the pricing performance of different asset pricing models. This allows one to abstract from an arbitrary choice of a single or a pair of characteristics to construct univariate or double-sorted portfolios as test assets. This level of abstraction and the explicit mapping between asset characteristics, loadings, and associated latent factors may be particularly suitable within the context of cryptocurrency markets, where the nature of the stochastic discount factor is yet arguably unknown. Put it differently, the IPCA efficiently aggregates a large set of asset characteristics that potentially could all be informative, given the relatively unexplored nature of the cryptocurrency market, and then allow the data to dictate if and how these characteristics can provide a reliable risk-based explanation of cryptocurrency returns.

Empirically, we investigate the returns dynamics of a large cross section of cryptocurrency pairs traded daily against the U.S. Dollar from September 1st, 2017 to September 1st, 2022 on more than 80 centralised exchanges. For each pair, we construct a set of 28 characteristics, which can be broadly categorized as on-chain activity, trading frictions, and past performances. On-chain activity contains measures of network growth, value, and adoption (see, e.g., Pagnotta and Buraschi, 2018; Cong et al., 2021b). We add a residual category dubbed as "other", that contains the equivalent of a CAPM alpha and a simple non-parametric downside risk measure. Our analysis shows that, by leveraging information in observable cryptocurrency characteristics to estimate latent factors and the corresponding dynamic loadings, researchers can better understand the risk-reward trade-off in cryptocurrency markets compared to the insights gained from traditional static latent factor models and long-short observable portfolios.

Our main contribution is fourfold: first, we show that a parsimonious IPCA model with few latent factors can provide a more accurate risk-based explanation of both the realised returns variation – i.e., systematic risks – and the difference in average returns – i.e., risk compensation –, when compared against standard latent or observable risk factors. For instance, a four-factor IPCA model (henceforth `IPCA4`) produces an out-of-sample total R^2 for individual cryptocurrency daily returns of 11.5%. For comparison, a benchmark factor pricing model with seven observable factors (henceforth `FF7`) – market, size, momentum, volatility, liquidity, past maximum daily returns, and network-to-market value – produces an out-of-sample R^2_{total} of 8%. In addition, the `IPCA4` provides a more accurate description of risk compensation in cryptocurrency markets, as highlighted by a positive out-of-sample predictive R^2 equal to 0.30%, against 0.03% and 0.02% obtained from a static principal component regression with seven latent factors (henceforth `PCA7`) or the `FF7` model, respectively.

The IPCA's success in explaining jointly the variation in both realised and expected returns come solely from the exposure to common risk factors and does not depend on mispricing effects, i.e., the intercept coefficients are restricted to zero for all assets. Yet, the gap in favour of the `IPCA4` increases when characteristic-sorted portfolios are used as test assets: the out-of-sample $R^2_{total,x}$

($R^2_{pred,x}$) obtained from the `IPCA4` model is 52% (1.9%), whereas the total and predictive out-of-sample performance of the benchmark (`FF7`) model is 19% (0.95%). The performance gap in favour of the `IPCA4` with respect to `PCA7` also increases for managed portfolios, with the latter producing an $R^2_{total,x}$ ($R^2_{pred,x}$) equal to 17.7% (0.5%) out of sample, daily.

The second main result pertains to a set of alternative dimensions through which the pricing performance of the IPCA is compared to both latent and observable risk factors model. We consider the time series R^2_{ts}, the cross-sectional R^2_{cs}, and the "relative pricing error" (RPE). The latter is defined as the magnitude of the model's unexplained average returns, i.e., the ratio between the alphas and the historical average returns. When a model does not explain any systematic part of the returns variation, the RPE is at, or above, 100%. The results show that the ability of the `IPCA4` model to explain both the time-series and the cross-sectional variation of the returns is substantially higher than that of the observable `FF7` model. For instance, the out-of-sample time-series R^2_{ts} of the `IPCA4` is 20.4% (46%) for individual assets (managed portfolios); this compares to a 18.8% (5.02%) obtained from the `FF7` model. In addition, while the `IPCA4` delivers an out-of-sample RPE of 10.9%, the competing `FF7` (`PCA7`) model generate a much higher relative pricing error of 63% (74%). This suggests that, as far as the pricing performance is concerned, the IPCA seems to provide a more accurate risk-based explanation of both the realised and expected cryptocurrency returns variation.

As highlighted by Kelly et al. (2019), the dual implication of IPCA's superior performance is that the IPCA latent factors are more consistent with mean-variance efficiency. We show that the four-factor IPCA specification achieves an ex-ante Sharpe ratio of 0.83, versus 0.18 (0.40) for the benchmark `FF7` (`PCA7`) model. Finally, the superior asset pricing performance of the IPCA compared to conventional models is confirmed by looking at the unconditional average absolute alphas across managed portfolios: the average absolute alphas of the 28 portfolios sorted on different asset characteristics is 0.17% daily for the `IPCA4` compared to a 0.59% from the `PCA7` and 0.57% for the `FF7`.

Our third main result relates to the dynamics of the IPCA factor loadings and the factors interpretation. We build upon Kelly et al. (2019) and Kelly et al. (2022), and test the significance of the characteristics used to discipline the betas through a semi-parametric bootstrap procedure. Our testing results show that, based on the baseline `IPCA4` model, the factor loadings, and therefore the conditional expected returns, are primarily driven by a handful of individual asset characteristics including liquidity, past performances, and volatility. The fact that only a small set of individual asset characteristics (7 of 28) is significant, coupled with the zero-alpha restriction in the baseline IPCA, suggests that these characteristics do not represent spurious compensation in the absence of risks.

In addition, our main results show that the latent IPCA factors are not spanned by observable long-short portfolios. This is directly tested both by leveraging the flexibility of the IPCA estimation methodology, and by a series of factor spanning regressions. A series of individual regressions of each latent factor on managed portfolios provides some interesting interpretations of the IPCA model. For instance, the first factor primarily correlates with the market beta and network growth measures, the second factor with the maximum daily return in the previous week, the third factor with trading volume, size and turnover, and the fourth factor is a combination of past performance measures, liquidity and volatility. This suggests that a combination of trading frictions and network growth might be the key determinant of risk premiums within the cryptocurrency space (see, e.g., Pagnotta and Buraschi, 2018; Cong et al., 2021b; Babiak et al., 2022).

The final main result of the paper relates to the alleged segmentation between cryptocurrency and traditional equity markets (see, e.g., Liu and Tsyvinski, 2021). The conventional wisdom posits that although cryptocurrency and equity markets are fundamentally segmented, the correlation between the two steadily increased since the outbreak of the Covid 19 pandemic. In this respect, our assessment asks: how relevant is the pricing information contained in standard equity portfolios for the cross section of cryptocurrency returns and vice-versa? The IPCA framework is particularly suitable because it allows us to directly test for the incremental explanatory power of equity risk factors while at the same time being agnostic on the nature of the common cryptocurrency factors. This allows us to understand the issue of fundamental correlations between cryptocurrency and equity risk factors in a self-contained framework.

The results suggest that once we control for IPCA latent factors, the information content of equity risk factors to explain the variation in realised and expected cryptocurrency returns is negligible. When we augment the IPCA fit on the cross section of individual cryptocurrency returns with the Fama and French (2015) five equity risk factors, the total and predictive R^2 remain virtually the same. In addition, a semi-parametric bootstrap test shows that none of the equity risk factors considered produce a statistical significant effect on the dynamics of realised or expected cryptocurrency returns, once we condition for the IPCA factors. Nevertheless, a series of factor spanning regressions shows that, albeit small, there is some correlation between the IPCA cryptocurrency factors and the Fama-French equity factors. For instance, the correlation between the first IPCA factor and the excess return on the equity market is highly significant and increased after March 2020.

Overall, the empirical results suggest that market segmentation may still potentially represent an impediment to cross-asset fundamental spillovers between equity and cryptocurrencies, compared to other asset classes such as bond (see, e.g., Kelly et al., 2022), foreign exchange and commodities (Asness et al., 2013). However, the presence of a moderate correlation between markets, as shown for instance by the first IPCA latent factor and the equity market portfolio, potentially suggests that investors' hopes on the "diversification" benefits of cryptocurrencies may have been ill-posed (see, e.g., Baek and Elbeck, 2015; Yermack, 2015; Biais et al., 2020).

The strong performance of the IPCA compared to traditional latent and observable factor models holds across different sub-samples, such as by slicing the cross section by quartiles on size, liquidity, and network growth, including when we divide the sample into pre- and post-COVID 19 outbreak, and using less granular weekly returns. The latter helps us to better understand the properties of the IPCA when the ratio between the number of assets and the number of observations increases, and provides some useful comparisons with the existing literature. In addition to Liu et al. (2022), who study the variation in weekly cryptocurrency returns in the pre-Covid period, based on a series of observable risk factors, our work is related to a growing literature that aims at understanding the trade-off between risks and rewards within the context of cryptocurrency markets, including Cong et al. (2021b), Makarov and Schoar (2020), Dobrynskaya (2021), Makarov and Schoar (2021) and Babiak et al. (2022), among others. It is also related to a large literature investigating how individual characteristics can be used to predict risk premiums (including Freyberger et al., 2020; Büchner and Kelly, 2022; Kelly et al., 2022, among others), and to a large literature studying the pricing dynamics and investment properties of digital assets (for example, Weber, 2016; Biais et al., 2020; Chiu and Koeppl, 2017; Cong and He, 2019; Cong et al., 2021a,c; Sockin and Xiong, 2020; Schilling and Uhlig, 2019; Abadi and Brunnermeier, 2018; Routledge and Zetlin-Jones, 2021).

2 Research design

2.1 Data description

We collect OHLC prices and trading volume from CryptoCompare.com and the data on on-chain activity from IntoTheBlock.com. The data are sampled daily from June 26th, 2014 to September 1st, 2022, with a day defined at a start time of 00:00:00 UTC. Daily prices and volume are aggregated across more than 80 different centralised exchanges which are deemed to provide a sufficiently reliable trading platform by CryptoCompare.[2] The aggregation across different exchanges is volume-weighted, that is, prices and trading volume are aggregated based on the exchange-specific trading activity. This procedure implies that larger and more established exchanges tend to have relatively more weight in the aggregation of the price and volume of a given pair compared to the weight of smaller/peripheral exchanges. All cryptocurrency pairs in the sample use USD as the quote currency, that is, USD represents the "domestic" currency in the sample (see, e.g., Liu et al., 2022). We only retain cryptocurrency pairs if they have all available data from CryptoCompare.com and IntoTheBlock.com after merging.

We introduce a variety of filters to mitigate the effect of erratic or suspicious trading activity. First, we exclude pairs that had zero traded volume or a zero price for any day t. Second, for each pair and day t we compute the ratio of traded volume to market capitalization and exclude pairs with a ratio greater than 1. This is a simple filter to screen out pairs which are potentially subject to "fake" volume or "wash trading" in a given day, meaning trading activity which is largely decoupled from the actual market value of a cryptocurrency on a specific date. The threshold is conservative because the median of the ratio is 0.01. Third, we screen out (1) all cryptocurrencies that are backed by or that track the price of gold or any precious metal, (2) so-called "wrapped" coins, e.g., Wrapped Bitcoin (WBTC), since they represent copies of existing tokens, (3) stablecoins, including those that are centralized (e.g., USDT, USDC) and algorithmically stabilized (e.g., DAI, UST) for all fiat currencies, and (4) coins that are synthetic derivatives, e.g., stETH, stSOL, as they track the value of an existing cryptocurrency. The screening is based on the classification provided by CoinMarketCap.com. Appendix A provides more details on the additional filters implemented in

[2]The exchanges that we include in the aggregation are the ones ranked from AA to B by CryptoCompare.com. The precise ranking of all exchanges appears on the company website at https://www.cryptocompare.com/exchanges//overview.

the aggregation step by CryptoCompare.com to mitigate the impact of fraudulent trading activity and/or malfunctioning API for individual exchanges. After applying all filters, we are left with an unbalanced panel of 395 cryptocurrency pairs.

Figure 1 overviews the data. The left panels compare the aggregated capitalization of cryptocurrencies in our sample with the total market capitalization. Two observations are noteworthy. First, the figure shows that our dataset covers a large fraction of the total market, from as much as 95% from the beginning of the sample until late 2019 and a still almost 70% at the end of the period under consideration. Although the cross section in our sample is smaller than the number of cryptocurrencies currently available – at the time of writing there are more than 19,000 different tokens according to CoinMarketCap.com – coverage of the market value is quite substantial.[3]

Second, there is considerable time variation in the market value of cryptocurrencies in our dataset. For instance, the sample includes the ICO mania of late 2017, the so-called "crypto-winter" of 2018-2019, the COVID-19 crash in March 2020 – which resulted in a 40% loss in Bitcoin (BTC) and even greater losses in the alternative coins –, and the subsequent boom and bust cycle that begun late 2021 and ended with the spark of the Ukraine war in early 2022. In addition, our sample includes major regulatory and institutional changes, including the ban by the Chinese government on crypto exchanges, the introduction of tradable Bitcoin and Ethereum futures contracts on the Chicago Mercantile Exchange (CME), the launch of the first traded Bitcoin Futures ETFs in October 2021, and the built-up of the transition of Ethereum from proof-of-work to a proof-of-stake protocol. In sum, our sample covers a variety of regulatory events and market scenarios.[4]

The right panels in Figure 1 report a snapshot of the time-series and cross-sectional dimensions of our panel of assets. A pair is included in the sample if it has been traded for at least 365 days, though the asset may not necessarily be available at the end of the period. As a result, our panel eliminates short-lived coins, which potentially represent scams, but includes failed coins with a relatively longer history of transactions. This helps to mitigate potential survivorship bias, and results in an unbalanced panel of cryptocurrencies. As shown in the right panels in Figure 1, there is a wide range for the length of the individual time series, with an upper bound of 3,000 observations. Also, the panels illustrate that the number of available cryptocurrencies is less than 50 before September 2017, rapidly increases

[3]Note that this concentration is a common feature of cryptocurrency markets (see, e.g., Babiak et al., 2022).

[4]Our sample of almost 400 pairs is rather consistent with the number of assets commonly listed on major exchanges. For instance, at the time of writing, Binance.com – the largest centralised exchange by trading volume – is listing 394 tokens as per Coinmarketcap.com.

to 395 towards the end of 2019, and then slightly decreases to almost 300 by the end of the sample. Although the data is available from 2014, our empirical analysis uses the panel of cryptocurrency returns starting from September 1st, 2017 due to the limited number of available assets before this date. Indeed, including the data before September 1st, 2017 could be problematic for the estimation of the IPCA model, as the number of characteristic-managed portfolios approaches the number of cryptocurrencies.

Nevertheless, we evaluate the impact of the size of the cross section on the main empirical results by performing three robustness checks. First, we replicate the empirical analysis for the sub-samples before and after the COVID-19 outbreak. This exercise demonstrates the performance of the IPCA model estimated based on a different sample size. It further challenges the IPCA framework in the environment with large price swings as observed in the cryptocurrency markets since early 2021. Second, we recursively re-estimate the IPCA model on an expanding window basis, starting from the initial 50% of the data. In this case, we evaluate the performance specifically in the second half and further assess the out-of-sample asset pricing properties of the IPCA fit. Third, we slice the cross section of assets in different quartiles based on size, liquidity and on-chain activity and shows that the IPCA systematically outperform the other competing factor analytics. Overall, the results of our robustness checks refute the concerns related to the size of the cross section, sampling issues, or the estimation procedure.

2.1.1 Characteristics. Table 1 provides an overview of 28 characteristics we use as instruments in the IPCA methodology. We group them into four categories: on-chain activity measures, including new addresses (`new add`) and network-to-market value (`bm`); trading frictions, such as the average daily bid-ask spread (`bidask`) and idiosyncratic volatility (`ivol`); past returns, such as momentum (`r22_1`) and short term reversal (`r2_1`); and others, such as the CAPM alpha (`capm` α) and the daily historical Value-at-Risk at 5% (`VaR(5%)`). We follow Freyberger et al. (2020) and Liu et al. (2022) in the classification of characteristics. Appendix B contains a detailed description of the characteristics, the construction, and the relevant references.

Table 2 reports summary statistics for various characteristics and return predictors. For each variable, we report the cross-sectional mean, median, standard deviation, and relevant percentiles of the cross-sectional distribution of the time-series averages. Notably, the distribution of most average characteristics is positively skewed. This is most evident for new addresses (`new add`) and active addresses (`active add`). Similarly, trading volume (`$volume`), market capitalization (`size`), and Amihud (2002) ratios (`illiq`) are highly positively skewed. This is due to a handful of large cryptocurrencies - Bitcoin and Ethereum, among others - that are more heavily traded and more liquid than the average cryptocurrency pair.

2.2 Observable factor portfolios

We first construct our own cryptocurrency market factor (`mkt`) as a the returns in excess of the risk-free rate on the value-weighted portfolio of the cryptocurrency pairs in our data. This produces a proxy for market risk that is best positioned to coincide with the variation in our cryptocurrency return panel. The risk-free rate is approximated as the daily one-month Treasury bill rate. The market factor is motivated by a relatively high concentration of cryptocurrency markets (see market capitalization in Table 2).

In addition to the market risk factor, we consider a variety of long-short portfolios following the existing empirical asset pricing literature (see, e.g., Liu et al., 2022). We construct a comprehensive list of long-short strategies based on size, momentum, volatility, liquidity, reversal, and on-chain activity. Some of these factors have been shown to capture a significant amount of the variation in realised and expected cryptocurrency returns (see, e.g., Brauneis et al., 2021; Cong et al., 2021b; Leirvik, 2021), while some others have been adapted from the equity literature (see, e.g., Fu, 2009; Nagel, 2012).

Our construction of observable factors follows a standard approach. Specifically, for each day we sort individual cryptocurrencies into quintiles based on the value of a given characteristic. We then combine the cryptocurrencies within each quintile into a value-weighted portfolio based on the current relative market capitalization of each pairs. The next day we track the return on each quintile portfolios. Notice for the momentum factors we consider a one day skipping period after the portfolio formation to mitigate the bid-ask bounce or short-term reversal effects (see, e.g., Nagel, 2012).

We calculate the returns on a long-short strategy as the spread between the returns of the fifth and first quintile portfolios, or the opposite, depending on the risk factor we investigate. Furthermore, because the market capitalization is highly skewed in few pairs (see Table 2), we apply a 40% restriction on the weight of a given pair within a given quintile portfolio (see, e.g., Jensen et al., 2022). The choice of 40% is inconsequential. However, it produces more realistic capital allocation within each portfolio, which could otherwise be almost uniquely concentrated in a single asset without considering the weight cap. Alternatively, we could employ an equal-weighted portfolio allocation scheme for each quintile. However, given the low liquidity of smaller cryptocurrencies, some of which are akin to microcap firms, this would mechanically inflate the returns of each long-short strategy (see, e.g., Babiak et al., 2022).[5]

We first construct a `size` factor by sorting digital assets based on their log market capitalization. We calculate the market capitalization of each pair as the current supply of coins times their current market price expressed in USD. The current supply is the number of coins or tokens that have been mined or generated and corresponds to the number that are currently in public and company hands, which are circulating in the market and/or locked/vested. We construct the risk factor as a long-short portfolio that goes long (short) on small (large) assets. We assume that shorting occurs on the margin at a 1x leverage ratio. As a result, each time the portfolio is rebalanced, one can invest only a fraction of wealth in new short positions.[6] Table 3 reports the average return and volatility of the `size` portfolio; consistent with Liu et al. (2022) the average returns on the size factor are negative.

In addition to market and `size`, we consider two alternative liquidity factors. First, for each trading day, we sort individual cryptocurrencies into quintile portfolios based on the value of their Amihud (2002) ratio. We calculate this as the ratio between the absolute daily return and the average daily trading volume in $mln. We construct the `illiq` risk factor as a long-short portfolio which goes long (short) on less liquid (more liquid) assets. We consider an alternative liquidity factor by replacing the Amihud (2002) ratio with the synthetic bid-ask spread measures as proposed by Corwin and Schultz (2012) and Abdi and Ranaldo (2017). We sort each asset in quintile based on their average bid-ask spread (see, e.g., Babiak et al., 2022), and construct the `bidask` risk factor by taking a long (short) position on value-weighted portfolios with the highest (lowest) bid-ask spread. Table 3 shows the sample performance of both portfolios. Interestingly, both risk factors produce large and negative Sharpe and Sortino ratios. Differently from the market portfolio, the `illiq` risk factor generates a positive returns skewness. This is not entirely unexpected from a long-short portfolio strategy vs long-only market allocation.

[5]In a set of unreported results we show that although the performance of equally-weighted portfolios grows stronger, the explanatory power of observable risk factors remain subpar the IPCA.

[6]Although quite complicated to implement, the equivalent of a short sale can be created via margin trading on major exchanges, including Binance, Poloniex, and Bitfinex. In practice, these exchanges offer the possibility to borrow a given crypto at the current market price, and to sell it, and then to buy it back later to cover the investor's position. Another interpretation one could give to our long-short portfolio is a weighting scheme relative to a benchmark, a value-weighted market portfolio. In this case, a long (short) position could be interpreted in relative terms as overweighting (underweighting) some cryptocurrency pair with respect to its market weight (see, e.g., Liu et al., 2022).

We also consider two alternative long-short portfolios based on either realised or idiosyncratic volatility. We compute the realised volatility factor (`rvol`) using the estimator proposed by Yang and Zhang (2000) based on OHLC daily prices with a rolling period of 30 days. We then sort cryptocurrency pairs into value-weighted quintile portfolios from low to high realized volatility. A short position is initiated in low-volatility pairs, whereas a long position is taken in high-volatility pairs. In addition, we follow Ang et al. (2006) and measure the idiosyncratic volatility for each cryptocurrency as the standard deviation of the residuals from a 30-day rolling window regression of the individual returns onto the market portfolio. The `ivol` strategy returns are the return differential between quintile portfolios of the lowest and highest idiosyncratic volatility. Table 3 shows that sorting pairs either by their realized or idiosyncratic volatility generates a negative and significant average return, with a Sharpe ratio comparable to both liquidity factors. This negative performance is consistent with previous literature (see, e.g., Liu et al., 2022).

Next, we consider a variety of alternative specifications for past performance. First, we consider a simple short-term reversal strategy (r2_1) as in Nagel (2012); Babiak et al. (2022). Then, we construct several cross-sectional momentum factors as introduced by Jegadeesh and Titman (2001). We consider three different "look-back" periods of $l = 7, 14, 22, 31$ trading days. We allocate each pair into a given quintile based on its cumulative log return over the previous l-days. We then construct a corresponding momentum strategy as the long-short portfolio that goes long (short) on past winner (loser) assets. Table 3 shows that, at least unconditionally, all momentum strategies produce positive Sharpe ratios, with an average value of 0.02 daily (0.4 annualised). In addition to momentum we also consider long-short portfolios constructed based on the maximum daily returns over the last $l = 7, 30$ days. We construct quintile portfolios from the lowest to highest maximum daily returns over the past l days. We thus construct a long-short strategy (`max`l) by taking a long position on the highest max and a short position on the lowest max portfolio. Consistent with the existing literature on equity markets both `max`l portfolios produce quite large negative Sharpe ratios (see, e.g., Bali et al., 2011).

The last three long-short portfolio strategies we consider focus on blockchain network activity. First, we consider an on-chain "value" strategy as inspired by Pagnotta and Buraschi (2018). We construct this "value" proxy by using the network-to-market value ratio (`bm`): the cumulative number of unique addresses over the current available supply, times the current USD price. As Pagnotta and Buraschi (2018) suggest, the intrinsic value of a cryptocurrency/token could be directly dependent on the growth of the network, which can be proxied by the cumulative number of blockchain addresses that actively transact in a native coin. By dividing the cumulative number of active addresses to the market capitalization, one can approximate the extent of market over- or under-valuation of a given token with respect to its network dispersion.

Clearly, in the absence of cash flows and a clear definition of book value, the `bm` measure represents an approximation, at best. Therefore, in addition to network-to-market value ratio, we follow Cong et al. (2021b) and also construct long-short portfolios by sorting assets based on the number of unique addresses that appeared for the first time in a transaction of the native coin in the network, or based on the number of unique addresses that were active on a given day in the network either as a sender or a receiver (see Appendix B). Table 3 shows that all risk factors based on on-chain activity produce a positive Sharpe ratio which is in line with the market portfolio.

2.3 Instrumented principal components analysis

A factor pricing approach is the most common empirical analysis to evaluating the trade-off between risks and rewards in financial markets. It assumes that the information content in the cross section of individual asset or portfolio returns can by summarised by a small set of factors. This approach does not depend on the asset class under investigation, and is grounded on fundamental asset pricing theory. Assuming the no-arbitrage condition holds, a stochastic discount factor m_{t+1} exists and the Euler equation $E_t\left[m_{t+1} r_{i,t+1}\right] = 0$ holds for any excess return $r_{i,t+1}$. Consequently, the conditional expected return satisfies

$$E_t\left[r_{i,t+1}\right] = \underbrace{\frac{Cov_t\left(m_{t+1}, r_{i,t+1}\right)}{Var_t\left(m_{t+1}\right)}}_{\beta_{i,t}} \underbrace{\left(-\frac{Var_t\left(m_{t+1}\right)}{E_t\left[m_{t+1}\right]}\right)}_{\lambda_t}, \tag{1}$$

in which $\beta_{i,t}$ is conditional exposure of asset i at time t to systematic risk factors and λ_t is the time-varying price of risks associated with factors. Assuming m_{t+1} is linear in factors f_{t+1}, the cross section of excess returns satisfies a linear factor model:

$$r_{i,t+1} = \alpha_{i,t} + \beta_{i,t} f_{t+1} + \epsilon_{i,t+1}, \tag{2}$$

where $E_t\left[\epsilon_{i,t+1}\right] = 0$, $E_t\left[f_{t+1}\epsilon_{i,t+1}\right] = 0$, $\lambda_t = E_t\left[f_{t+1}\right]$ and $\alpha_{i,t} = 0$ holds for all i and t.

Notably, the nature of factors f_{t+1} and the dynamics of intercepts $\alpha_{i,t}$ and loadings $\beta_{i,t}$ are left unspecified by the asset pricing theory. In this paper, we jointly estimate $\alpha_{i,t}, \beta_{i,t}$ and f_{t+1} via the instrumented principal components analysis (IPCA) method developed by Kelly et al. (2019) and used more recently by Büchner and Kelly (2022) and Kelly et al. (2022) for modelling option and corporate bond returns. The IPCA framework assumes that risk factors are latent and extracted from the cross section of test assets, whereas intercepts and loadings are time-varying and linear in asset characteristics:

$$\alpha_{i,t} = \boldsymbol{z}'_{i,t}\Gamma_\alpha, \qquad \beta_{i,t} = \boldsymbol{z}'_{i,t}\Gamma_\beta, \tag{3}$$

where $\boldsymbol{z}_{i,t}$ denotes an $L \times 1$ vector of (cryptocurrency) characteristics. The mapping between characteristics and dynamic intercepts and factor loadings is linear and is determined by the matrices $\Gamma = [\Gamma_\alpha, \Gamma_\beta]$. The main hypothesis throughout this paper is that the coefficients of the intercept Γ_α to be zero for all assets. This implies that the variation in the realised and expected returns is consistent with a beta/expected return model, rather than being a reflection of compensation without risk. By restricting $\Gamma_\alpha = 0$, the state-space in Eq.(2)-(3) is estimated via an alternating least squares approach, which iterates the first order conditions of Γ_β and f_{t+1}

$$\widehat{f}_{t+1} = \left(\widehat{\Gamma}'_\beta Z'_t Z_t \widehat{\Gamma}_\beta\right)^{-1} \widehat{\Gamma}'_\beta Z'_t r_{t+1}, \qquad \forall t, \tag{4}$$

and

$$\text{vec}\left(\widehat{\Gamma}_\beta\right) = \left(\sum_{t=1}^{T-1} Z'_t Z_t \otimes \widehat{f}_{t+1}\widehat{f}'_{t+1}\right)^{-1} \left(\sum_{t=1}^{T-1} \left[Z_t \otimes \widehat{f}'_{t+1}\right]' r_{t+1}\right), \tag{5}$$

where Z_t and r_{t+1} denote the stacked arrays of instruments and returns, respectively. In the main results we also estimate an unrestricted model with $\Gamma_\alpha \neq 0$ and then test for the null hypothesis that $\Gamma_\alpha = 0$ via non-parametric bootstrap. The extension of the alternating least squares approach for the unrestricted model becomes consequential by simply adding a constant to the vector of factors. One comment is in order. Table 2 shows that a significant fraction of characteristics have a rather skewed cross-sectional distribution. To mitigate the impact of skewed observations on the estimation procedure, we build upon Kelly et al. (2019) and cross-sectionally rank, demean, and scale the characteristics to exist in the $[-1,\ 1]$ interval. This places characteristics on the same investment scale as standard long-short portfolios, and at the same time mitigate the impact of outlying observations so that the estimate of Γ is directly comparable across characteristics.

2.3.1 Motivating evidence for IPCA. A natural starting point to understand the value of instrumenting the conditional betas within an otherwise standard factor pricing model is a panel predictive regression of cryptocurrency returns on the same set of characteristics used to estimate the IPCA. The left and middle panels of Table 4 reports two different sets of results; a simple pooled OLS where individual fixed effects have been discarded, and a panel regression with individual fixed effects. Notice that, to retain some sort of direct mapping with the IPCA estimation, we have cross-sectionally re-scaled the characteristics to be in the $[-1,\ 1]$ interval. We find that a moderately large set of characteristics is indeed significant, although the in-sample predictive R^2 is essentially zero. Even when we add fixed effects, the fit does not significantly improve. In other words, although there is evidence that characteristics and future returns tend to correlate, a standard panel predictive regression would not excel in explaining the dynamics of realised returns.

The right panel of Table 4 expands the evidence on the raw returns and investigates the extent to which cryptocurrency characteristics predict market betas. Instead of using the one-day ahead returns as the target variable in the panel predictive regression, we plug in a measure of the future realised market betas, calculated based on a 30-day rolling window regression. To avoid overlapping observations, we consider the 30-day market beta at time $t+30$ in a predictive regression of the form: $\beta^{mkt}_{i,t+30} = \gamma \boldsymbol{z}_{i,t} + \text{constant} + \text{error}$.

The results shows that characteristics do indeed significantly predict market betas with a panel R^2 of 1.9%. Compared to the 0.06% of the pooled OLS on the raw returns, the individual characteristics have a substantially higher predictive content when it comes to individual risk exposure vis-á-vis returns. Although this seems relatively low, recall that there is a fair degree of noisiness in the market betas estimated with only 30 days of realised returns. In addition, several of the characteristics are significant. Among the strongest predictors, we can find the `$volume` and `illiq` measures, the reversal factors (`max` and `rel_to_high`), idiosyncratic volatility (`vol`), and turnover (`turnover`), each of which with highly significant regression coefficients – coefficients which are directly comparable given the characteristics have been cross-sectionally ranked and standardised.

Taken together, the results in Table 4 suggest that cryptocurrency characteristics offer some predictive ability for the market betas. On the other hand, the predictive content for the raw returns seems to be relatively weaker. The specification of the panel regression is somewhat ad-hoc. The results provided in Table 4 cannot help to distinguish between mispricing or risks as the main driving force behind returns predictability (see, e.g., Daniel and Titman, 1998). The IPCA helps to answer this question by unifying both lines of inquiry. As highlighted by Kelly et al. (2022), the IPCA allows a researcher to distinguish the extent to which whether characteristics explain asset realised and expected returns due to a risk-based channel, or if characteristics predict returns because they proxy for alpha above and beyond factor risk (i.e., mispricing or "anomaly" channel). The next section pursues this investigation.

3 Empirical results

We report a series of statistical measures to assess the asset pricing performance of the IPCA compared to both static and conditional observable factor models, and to a standard PCA. We follow Kelly et al. (2019) and compute two alternative measures of aggregate goodness-of-fit for the panel of cryptocurrency returns, namely the total R^2_{tot} and the predictive R^2_{pred}, which are defined as

$$R^2_{total} = 1 - \frac{\sum_{i,t}\left(r_{i,t+1} - \widehat{\beta}'_{i,t}\widehat{f}_{t+1}\right)^2}{\sum_{i,t} r^2_{i,t+1}}, \qquad R^2_{pred} = 1 - \frac{\sum_{i,t}\left(r_{i,t+1} - \widehat{\beta}'_{i,t}\widehat{\lambda}\right)^2}{\sum_{i,t} r^2_{i,t+1}}, \tag{6}$$

with $\widehat{\lambda}$ denoting the unconditional time-series mean of the factors. The R^2_{tot} quantifies the extent to which a given factor model captures the total variation in the realised returns. Instead, the R^2_{pred} captures how well differences in the average returns are explained through the model's ability to describe risk.[7] We also compare different latent and observable factor model specifications based on two additional metrics used in asset pricing literature (see, e.g., Büchner and Kelly, 2022; Kelly et al., 2022). More precisely, we first construct a time series R^2_{ts} by aggregating individual $R^2_i = \left(1 - \frac{\sum_t\left(r_{i,t+1} - \widehat{\beta}'_{i,t}\widehat{f}_{t+1}\right)^2}{\sum_t r^2_{i,t+1}}\right)$ across the time series

$$R^2_{ts} = \frac{1}{\sum_i \tau_i}\sum_i R^2_i, \tag{7}$$

[7]Notice that the denominator represents the square of the returns not demeaned. However, this is because the historical average daily returns of individual assets is close to zero statistically speaking, that is a forecasting from the unconditional mean would not deviation significantly from a forecast at zero. This is consistent with Gu et al. (2020) who argue that out-of-sample comparison of fits against historical mean is flawed when it comes to individual assets.

with τ_i the number of non-missing observations in a given cryptocurrency pair. Second, we report the average over the time series the cross-sectional $R_t^2 = \left(1 - \frac{\sum_i \left(r_{i,t} - \widehat{\beta}'_{i,t_1} \widehat{f}_t\right)^2}{\sum_i r_{i,t}^2}\right)$,

$$R_{cs}^2 = \frac{1}{T} \sum_t R_t^2, \tag{8}$$

where R_{cs}^2 quantifies the cross-sectional strength of the signal produced by either the IPCA or the observable factor models. This is akin to a Fama and MacBeth (1973) where the factors are "tradable", or represents replicable trading strategies. The last metric used to compare different models is the so-called *relative pricing error* (RPE), which is defined as

$$\text{RPE} = \frac{\sum_i \alpha_i^2}{\sum_i \overline{r}_i^2}, \tag{9}$$

with $\overline{r}_i = \tau_i^{-1} \sum_t r_{i,t+1}$ the asset's time series average returns and $\alpha_i = \tau_i^{-1} \sum_t \left(r_{i,t+1} - \widehat{\beta}'_{i,t} \widehat{f}_{t+1}\right)$ the average time series error, or "pricing error". A value of RPE closer to zero implies that a given model explain most of the systematic variation in the returns, whereas an RPE at, or above, 100%, implies that a given model does not capture any, or very little, systematic risk. Notice that in the main results we focus on the relative pricing error for the managed portfolios only. This is because both the alphas and the unconditional average returns for individual assets are all very close to zero. As a result, the RPE from daily individual returns would be very sensitive to outlying observations. On the other hand the larger scale and more compact distribution of alphas and average returns on the managed portfolios allows for a more reliable measurement of the RPE as a ratio.[8]

3.1 Asset pricing performance

3.1.1 In-sample results. Panel A of Table 5 reports the performance of a restricted IPCA model with $\Gamma_\alpha = 0$. This implies that the variation of cryptocurrency returns are consistent with a beta/expected return model. In addition to the performance for the restricted IPCA model, Table 5 also reports the bootstrap p-values for the hypothesis test of $\mathcal{H}_0 : \Gamma_\alpha = 0$ with a number of latent factors ranging from $K = 1$ to $K = 7$.[9] For $K = 1, 2, 3$ the bootstrap results show that we

[8] Although the RPE for individual assets is much more noisy, the results still highlight a comparatively lower relative pricing error from the IPCA against PCA and observable risk factors. Results are available upon request.

[9] We follow Kelly et al. (2019) and for each model specification, we construct the test statistic based on the identical implementation of a "wild residual" bootstrap approach. We first draw 10,000 pseudo-samples under the null hypothesis $\mathcal{H}_0 : \Gamma_\alpha = 0$. For each sample, we construct a Wald-type statistic measuring the distance between the restricted and unrestricted models. We then calculate the fraction of simulated statistics exceeding the corresponding value from the

can strongly reject the null hypothesis $\Gamma_\alpha = 0$. However, the bootstrap p-values show that when the number of latent factors is $K > 3$, the variation in cryptocurrency returns is solely captured by compensation to risk. Furthermore, the information criterion (IC) calculated as in Bai and Ng (2002) shows that a four-factor IPCA model maximizes the information content of the latent factors. By coupling the bootstrap and the IC, we choose the `IPCA4` as our *baseline* IPCA specification. This choice is further supported by comparing the performance of the restricted and unrestricted IPCA models. For the interested reader, the full comparison between restricted and unrestricted IPCAs is reported in Appendix C. For $K > 3$ factors, the additional variation in expected returns that is captured by an unrestricted intercept is negligible compared to a restricted IPCA model.

In addition to the IPCA, Panel A also reports the results from a standard, static, PCA. Given the unbalanced nature of the panel of returns and the presence of missing data, the PCA is implemented by using an alternating least squares procedure to estimate the static principal components (see Ilin and Raiko, 2010). At the individual returns level, the R^2_{tot}, R^2_{ts} and R^2_{cs} from a PCA with seven latent factors (`PCA7`) are comparable, if not better, than our baseline `IPCA4`. However, PCA provides a dismal description of expected returns at the cryptocurrency pair level; the R^2_{pred} is an order of magnitude lower than the IPCA for each K. This suggests that although a static PCA provides a fairly accurate description of the common variation in the returns, it is the information from the characteristics and the consequential dynamics of the loadings that provides a better description of risk compensation across assets.

Panel A also reports the same set of performance measures using the set of characteristic-based portfolios as test assets. As a result, all R^2 measures can be redefined in terms of managed portfolios based on IPCA parameters Γ_β (see Kelly et al., 2019).[10] The explanatory power of the IPCA is markedly stronger for portfolio returns than for returns on individual assets. For instance, the baseline `IPCA4` specification generates an $R^2_{tot,x}$ ($R^2_{pred,x}$) of 51.9% (1.93%). This compares to 16.9% (0.37%) for the static `PCA7` model. Both the time-series and cross-sectional aggregations of the R^2 are also in favour of the `IPCA4` model: the $R^2_{ts,x}$ ($R^2_{cs,x}$) for IPCA is equal to 45% compared to 10% (43% compared to 13%) from the `PCA7` model. The RPE is also significantly in favour of the IPCA model;

data to obtain the p-value for the IPCA model considered.

[10] Kelly et al. (2019) argues that the IPCA methodology incorporates a portfolio notion that circumvents a common problem in empirical asset pricing; the choice of the test assets through which a given factor model can be tested. When looking at equity markets, researchers tend to use double-sorted portfolios formed on different characteristics, including size and book-to-market ratios (see, e.g., Fama and French, 2015). Nevertheless, the choice of the most appropriate portfolios has been a source of debate (Lewellen et al., 2010; Daniel et al., 2012).

the RPE for the IPCA4 is down to 6.6% compared to a 78% from the static model with seven latent factors. This finding is consistent with prior results for equity, corporate bond, and option returns, where the relative price error of IPCA tends to be smaller for managed portfolios (see, e.g., Kelly et al., 2022).

Turning to observable risk factor models, Panel B of Table 5 provides the same set of performance metrics obtained by replacing the latent factors, either with or without instruments, with a series of long-short portfolio returns. Liu et al. (2022) shows that three observable risk factors, the excess returns on the market, size, and momentum, can span a great deal of the cross-sectional variation cryptocurrency returns at a weekly frequency. Cong et al. (2021b) consider a four factor based on the network-to-market value ratio to proxy for a valuation ratio based on on-chain network activity. We expand their factor structure based on the results in Table 3: in addition to `mkt`, `size`, `r22_1` and `bm`, we include liquidity (`bidask`), realised volatility (`rvol`), and reversal (`max7`). In total we consider seven observable factors, so that a comparable analysis can be made with the IPCA and the PCA specifications. A full description of each factor portfolios is provided in Section 2.2.[11]

The static factor model specification follows a standard factor pricing model with the betas estimated from a panel regression of cryptocurrency returns on observable risk factors. The instrumented version of the observable risk factor model can still be estimated using the IPCA procedure. More specifically, letting g_t denote the set of observable risk factors, the instrumented principal component model can be rewritten as

$$r_{t+1} = \boldsymbol{z}_{i,t}' \Gamma_\beta g_{t+1} + \eta_{t+1} = \text{vec}\left(\Gamma_\beta\right)' \left(\boldsymbol{z}_{i,t} \otimes g_{t+1}\right) + \eta_{i,t+1}, \tag{10}$$

Given that the factors are pre-specified, this specification can be estimated by evaluating only the matrix of loadings Γ_β from the associated first-order condition. We impose a zero-intercept constraint, i.e., $\Gamma_\alpha = 0$, for both the static and dynamic observable factor models to align with the baseline IPCA specification and isolate the jointly explanatory power of the latent factors and the individual characteristics. The notation is consistent across models, meaning that with `FF`l (`IFF`l) we indicate a static (instrumented) observable factor model with $l = 1, \ldots, 7$ risk factors included.

The results have three interesting aspects. First, the explanatory power of the IPCA factors

[11] In a set of unreported results, we replace `bidask` with `illiq`, `rvol` with `ivol`, and `max7` with `max30`. All alternative specifications produce lower R^2_{tot} and R^2_{pred}, so we choose the best possible specification for the observable factors model.

outperforms observable risk factors by a significant margin. For instance, our baseline `IPCA4` model generates a 13.4% R^2_{tot}, whereas a static factor model with seven portfolios (`FF7`) delivers a 10.8% R^2_{tot}. Perhaps more importantly, a model with seven observable factors (`FF7`) produces a dismal explanation of the conditional expected returns, with an R^2_{pred} of -0.02% compared to 0.32% from the `IPCA4`. Second, turning to the time-series and cross-sectional metrics, while the R^2_{ts} is somewhat comparable between the `IPCA4` and the `FF7` model, the R^2_{cs} from the former is 11.3% versus a 8.6% static observable factor model with seven risk factors. Third, the performance gap in favour of the IPCA is markedly larger for portfolio returns. For instance, the baseline restricted `IPCA4` specification generates an $R^2_{tot,x}$ ($R^2_{pred,x}$) of 51.9% (1.93%) against a 19.5% (0.8%) from the static and 21.8% (0.88%) from the instrumented observable factor models.

Both the time-series and cross-sectional aggregation of the R^2 for managed portfolios are also in favour of the `IPCA4` model: the $R^2_{ts,x}$ is equal to 45% for `IPCA4` versus 12.3% from the `FF7` model. Similarly, the $R^2_{cs,x}$ for the `IPCA4` is equal to 43% compared to 12.9% for the `FF7` model. The relative pricing error is also considerably in favour of the IPCA, with an RPE of 6.66% for the `IPCA4` against a 63.5% for the `FF7` model. Interestingly, the performance gap between a static versus a dynamic observable factor model is quite small.

3.1.2 Out-of-sample results. We expand the previous in-sample performance analysis and conduct an out-of-sample evaluation of the IPCA and other competing factor model specifications. The models are estimated for an expanding window starting from March 1st 2020, that is, the first half of the data available are used as "burn-in" sample. We perform forecasts for each period, based on the estimated parameters and factor returns at that time. For observable factors, we use the actual portfolio returns in the forecast construction. For the IPCA factors, our computations of the out-of-sample factor returns follows the framework of Kelly et al. (2019). We then evaluate the out-of-sample performance of each model based on the realized returns and model forecasts of individual cryptocurrencies and managed portfolios.

Table 6 summarises the results. The specification with four latent factors explains 11.5% of the total variation in the out-of-sample realized individual returns. By comparison, the R^2_{tot} obtained from the `FF7` and `IFF7` observable factor models is 8% and 8.5%, respectively. Further, IPCA produces a substantially better risk-based explanation of the average returns across individual assets: the out-of-sample R^2_{pred} is 0.3%, compared to a 0.03%, 0.02%, and 0.09% obtained from the `PCA7`, the static

FF7 and the instrumented IFF7 observable factor models, respectively. The R^2_{ts} and the R^2_{cs} are also higher for IPCA against the observable factor models, regardless of a number of factors considered.

Similar to the in-sample results, the gap between the IPCA and observable factor models widens by using managed portfolios as test assets. The $R^2_{tot,x}$ for the IPCA4 increases to 52.7%, while it reaches the maximum of 19.2% (22.23%) for the FF7 (IFF7) model. Perhaps more importantly, the IPCA4 substantially improves upon the PCA7, and both observable factor model specifications, i.e., FF7 and IFF7, when it comes to describe risk compensation. This is shown by a substantially larger $R^2_{pred,x}$ – 2.2% for IPCA4 versus 1.1% for the best performing competing factor model – and a much lower relative pricing error RPE – 10.9% for the IPCA4 versus a 57.3% obtained from the best performing alternative factor model. This result confirms that, compared to both PCA and observable factors, the IPCA better captures some fundamental risk-reward relationships that would be otherwise buried in the noise of highly volatile daily returns. More generally, regardless of the number of factors required to eliminate the mispricing, the IPCA factors maintain a fine statistical performance throughout.

3.1.3 Mean-variance efficiency. To test the mean-variance efficiency of the IPCA factors vs static latent/observable risk factors, we carry out a series of simple asset pricing tests based on different portfolios as test assets. We study the mean-variance efficiency of two sets of portfolios – a set of managed portfolios produced using the IPCA methodology, and a series of double-sorted portfolios based on size and alternative characteristics. We compare the baseline IPCA4 model against the PCA7 and the instrumented observable factor model IFF7.

Figure 2 reports the results. For convenience, we highlight significant alphas with filled markers. The plots also report the average absolute alpha for each specification, to quantify the average size of mispricing across different models. Note that, for the conditional factor models IPCA4 and IFF7, the alphas are computed as the time-series average of the period-by-period portfolios residuals. For the static factor model PCA7, the alphas are computed as intercepts from time series regressions of portfolio returns on the latent factors. For comparability, we assume a portfolio volatility target of 5% daily, consistent with the historical volatility of long-short portfolios, and rescale the portfolio weights accordingly using only backward looking information.

The main results confirm that the IPCA significantly reduces mispricing: the average absolute pricing error for the IPCA4 is 0.16% on a daily basis, compared to 0.59% and 0.49% from the PCA7, IFF7 models, respectively. Thus, IPCA produces less than a half average absolute pricing errors than do benchmark latent factor models and existing portfolios. We also find that the estimated alphas from both the PCA and the observable factors are clustered around the 45-degree line. This indicates that static latent or observable factors may not provide an accurate risk-based explanation of managed portfolio returns.

Table 7 breaks down the alphas from the IPCA and both the PCA and observable factor models for each characteristic separately. Panel A reports the results for the full sample of daily returns corresponding to Figure 2. This analysis shows that the average absolute alpha from IPCA is substantially smaller than those in the competing factor models. Further, the alphas of managed portfolios are uniformly smaller in absolute value in the conditional IPCA model, with the exception corresponding to the market beta (`capm` β). More generally, the IPCA seems to unequivocally provides a more accurate risk-based description of the cross-sectional variation of the managed portfolio returns compared to both PCA and observable factor models.

We investigate the mean-variance efficiency of IPCA factors also using a set of 25 double-sorted portfolios as test assets. These portfolios should provide a more challenging case because they are not targeted by the IPCA estimation. Figure 3 reports the alphas from `IPCA4`, `PCA7`, or `IFF7`, respectively. We report the results for portfolios sorted on `size` and `r22_1` (Panel A), `size` and `bm` (Panel B), and `size` and `max7` (Panel C). Two observations are noteworthy. First, double-sorted portfolios represent indeed a more challenging set of test assets for IPCA. For instance, the average absolute daily alpha for portfolios sorted on `size` and `bm` is 0.24% versus the 0.16% on the 28 managed portfolios. This is also reflected in the performance of PCA. With the exception of the double-sorted portfolios on `size` and `max7`, the `PCA7` and `IPCA7` produces a similar average absolute pricing error. Yet, the pricing performance of the `IFF7` is rather dismal compared to the IPCA. For instance, the average absolute alpha for the `size` and `r22_1` portfolios is 0.3% versus 0.18% from the `IPCA4`. A similar gap applies to the `size` and `bm` double-sorted portfolios.

The second observation from Figure 3 pertains the significance of the alphas and the correlation with the raw return across portfolios. Take for instance the 25 portfolios sorted on `size` and `r22_1`; 18 (12) portfolios have significant alphas in the `IFF` (`PCA7`) model, whereas the `IPCA4` produces at most 5 portfolios with significant abnormal returns. The discrepancies in the significance of the pricing errors between the IPCA and the competing strategies persists across different sorting characterstics as shown in Panel B and C of Figure 3. In addition, when we visually assess the distribution of alphas, we find that portfolio alphas from the PCA and observable factor models exhibit a clear pattern: portfolio alphas increase with raw returns. This suggests a systematic shortcoming of both the PCA and observable risk factors in pricing the double-sorted test portfolios.

3.1.4 Factor tangency portfolios. We now compare the IPCA, PCA, and observable factors, and report the ex ante unconditional tangency portfolio performance for each group of factors, to describe their multivariate efficiency. The optimal allocation is based on recursive forecasts that we carry out by expanding the window of observations starting from March 1st 2020. We then calculate the mean-variance portfolio using the mean and covariance matrix of estimated factors through t and tracking the post-formation $t+1$ return. We assume a portfolio volatility target of 5% daily, consistent with the historical volatility of long-short portfolios, and rescale the portfolio weights accordingly using only backward looking information.

Table 8 shows summary statistics of tangency portfolios combining up to seven IPCA or observable factors. It reports the daily average returns, the Sharpe ratio and skewness, as well as the alphas of tangency portfolios from both a CAPM model (α_{CAPM}) and the alpha obtained by regressing the returns of the tangency portfolios on the seven observable risk factors considered in the main analysis. The tangency portfolio for the IPCA model yields a daily Sharpe ratio of 0.84, which is twice as large as the best Sharpe ratio of 0.4 for the `FF7` model. When we consider more than four latent factors, the performance of the IPCA tangency portfolios do not significantly improve. This is consistent with the bootstrap results in Table 5, which suggests that a four factor IPCA specification is sufficient to provide a risk-based explanation of cryptocurrency returns. The tangency portfolio constructed from PCA do not outperform the equivalent based on observable factors. For instance, the daily Sharpe ratio from `PCA7` is half the `FF7` equivalent.

On a risk-adjusted perspective, the tangency portfolios from the IPCA substantially outperform both PCA and observable risk factors. For instance, the α_{CAPM} for the `IPCA4` is as high as 4%, against a 1.1% and 1.7% obtained from the `PCA7` and `FF7`, respectively. Interestingly, the seven observable factors can not span the performance of the tangency portfolios based on the `IPCA4` or `PCA7` factors. The α_{F7} is 3.7% (t-stat = 19.5) and 1.05% (t-stat = 3.9), respectively. The spanning property is confirmed by looking at the tangency portfolios from the observable factors: once conditioning on the factors itself, the alphas are all economically negligible and not statistically significant.

3.2 Sub-sample analysis

Figure 1 highlights two main aspects of the sample under investigation, and in fact, of the cryptocurrency market at large. First, the total market capitalization significantly increases from the onset of the COVID-19 crisis; from the roughly \$300 billions in March 2020, the total market value increased tenfold, to an astonishing \$3 trillions towards the end of 2021, then lost more than 60% of its value by the end of the sample. Second, the size of the cross-section is fairly unbalanced before March 2020, but stabilizes somewhat thereafter. This is a general feature of the cryptocurrency market, with the number of tokens at the end of the sample almost doubling from the approximately 10,000 tokens available in early 2020.

To test the reliability of the IPCA framework in different scenarios and market conditions, in relation to both time-series and cross-sectional variation, we replicate the main empirical analysis for two different non-overlapping sample: a sub-sample from September 1st 2017 to March 1st 2020 (the solid red vertical line in Figure 1), and another sub-sample from March 2nd 2020 to September 1st 2022. It is worth reiterating that by dividing the sample, we challenge the IPCA along two dimensions. For the first sub-sample the panel of cryptocurrency pairs is highly unbalanced. It is smaller on average, and steadily increases over time. For the second sub-sample, the cryptocurrency market experienced significant drawdowns and volatility, while the size of the cross-section remained relatively more stable. Such abrupt variations should provide additional insights into the robustness of the asset pricing results across different conditions.

Table 9 reports the results for both sub-samples. Both the bootstrap test and the Bai and Ng (2002) information criteria support a four factor structure for the IPCA across different periods. Similarly, there is no sensible reduction of the relative pricing error RPE beyond three latent factors. The performance of the IPCA versus PCA and observable risk factors is quite heterogeneous across sub-samples. For instance, when we compare the two sub-samples, the R^2_{pred} for `IPCA4` improves from 0.29% pre-Covid to 0.4% in the second sub-sample. Furthermore, the PCA and the observable factor models both provide a less accurate description of the risk compensation across individual cryptocurrency pairs with an R^2_{pred} of 0.17% and 0.2%, respectively, in the period post March 1st 2020. However, the PCA shows a larger R^2_{tot} for individual assets compared to the IPCA, and both substantially outperform a `FF7` model.

The R^2_{ts} and the R^2_{cs} also provide some mixed results, with the PCA substantially improving upon the IPCA over the second sub-sample. This is possible due to the more balanced nature of the panel of observations over the second period. Nevertheless, compared to observable risk factors, the IPCA model provides a better fit for the variation in both realized and expected returns as well as a better characterization of both the time series and cross-sectional variation in the returns. This is particularly evident for risk compensation as approximated by the R^2_{pred}.

The use of managed portfolios as test assets provides some more clear-cut evidence in favour of the IPCA. For instance, the $R^2_{tot,x}$ is 51% for the `IPCA4` versus 21% and 24.9% from the `PCA7` and the `FF7` models, respectively, in the period after March 1st 2020. The same applies for the $R^2_{pred,x}$, with the observable risk factors that deliver a 2.11% versus a 3.33% obtained from the `IPCA4`. Perhaps more importantly, the IPCA provides a substantially lower relative pricing error compared to both competing classes of factor models; the RPE from the `IPCA4` is 8.4% (6.9%) for the first (second) sub-sample, compared to a 90% and 84% (63% and 43%) from the `PCA7` and `FF7`, respectively, over the first (second) sub-sample.

3.2.1 Mean-variance efficiency. We replicate the asset pricing tests as reported in Section 3.1.3. Figure 4 reports the scatter plot of the alphas obtained from the `IPCA4`, the `PCA7` and the `IFF7` model, respectively. As for the full sample, for convenience, we highlight significant alphas with filled markers. The plots also report the average absolute alpha for each specification, to quantify the average size of mispricing across different models. The main results confirm that the IPCA pricing performance is more consistent with mean-variance efficiency.

Turning to the first sub-sample, the average absolute pricing error across managed portfolios is 0.18%, compared to 0.51% from the `PCA7` and 0.47% from the `IFF7` factor models. Panel B shows that the gap in terms of mean-variance efficiency widens in favour of the IPCA over the post-March 1st 2020 period. For instance, the average absolute alpha from the `IPCA` is at 0.22% daily, against a 0.77% and 0.54% obtained from the `PCA7` and `IFF7`, respectively. Interestingly, the significance of individual alphas is slightly higher over the second sub-sample. For instance, pre-March 2020 there are 7 managed portfolios alphas from the IPCA that are significantly different from zero. Instead, over the post-March 2020 period the number of significant alphas from the IPCA grows to 9.

Figure 4 also shows that the estimated alphas from the PCA and the observable factors are more clustered – compared to the IPCA – around the 45-degree line. This holds across both sub-samples. Therefore, the evidence suggests that conventional latent or observable factor models may not be able to provide an accurate risk-based explanation of the cross-sectional variation of managed portfolio returns over time. To a large extent, the results in Table 9 and Figure 4 confirm that the IPCA model substantially reduces mispricing.

3.3 Individual assets quality and IPCA performance

Intuitively, returns on "low-quality" pairs, namely smaller cryptocurrency pairs with high trading frictions, tend to exhibit different behavior in terms of their covariances with characteristics, including liquidity and downside risk. For instance, Babiak et al. (2022) show that liquidity risk within the context of cryptocurrency markets is not uniformly spread across assets, but tends to be concentrated on assets with smaller market capitalization and lower trading volume. More generally, there is abundant evidence in the equity literature that adverse selection and information asymmetries tend to be concentrated on smaller, high volatile assets (see, e.g., Easley and O'Hara, 2004; Hendershott and Seasholes, 2007). This cross-sectional heterogeneity possibly could potentially affect the performance of factor pricing models.

In particular, the fact that smaller assets tend to be less liquid and more volatile raises the question of whether adding smaller assets actually has any significant statistical and economic effect on the main results, only because factor models may capture the variation of smaller assets at the expenses of larger ones. It is worth noting that the concept of "small" vs "large" assets in the context of cryptocurrency markets has non-trivial implications, considering the evident market concentration and skewed distribution both in size and trading activity (see Table 2). In practice, any asset with a market capitalization below the top 150, which at the time of writing is roughly \$150 millions, can be considered a micro-cap by equity standards.

To better understand the role of lower quality pairs on the asset pricing performance of the IPCA versus both PCA and observable factor models, we break out model R^2's for individual returns grouped according to three different characteristics. Each day, we sort the cross section of individual assets in quartiles based on market capitalization, number of active addresses or the average daily trading volume (see Section 2.2 for details). Then we measure the R^2_{tot} and R^2_{pred} from the IPCA, the PCA and the observable risk factors separately for each quartile. Note these results are not based on separate model re-estimation for each group of assets. This would mechanically allow each factor model to fit different groups based on different parameters or weighting schemes. Instead, we slice the factor model performance by keeping factors and parameters fixed at their estimates from the unified sample, and we recalculate the R^2's among each group of assets.

Table 10 reports the results. When we focus on the IPCA performance, the ability to explain the common variation in the realised returns increases with the market capitalization, the number of active addresses and the average trading volume. For instance, the R^2_{tot} from the `IPCA4` for smaller assets is 9%, whereas is 27% for the large assets. Similarly, the explained total variation from the `IPCA4` is 11% for cryptocurrencies with a low number of active addresses, while is 22% for assets with a higher network activity. Higher trading volume also coincides with a higher R^2_{tot} from the `IPCA4`; 8.34% for assets with low trading volume compared to 28.7% for assets with high trading volume.

Opposite to the R^2_{tot}, the ability of the IPCA to describe the differences in the expected returns across assets seem to be inversely related to assets quality. For instance, the R^2_{pred} from the `IPCA4` is 0.48% for the smaller assets, whereas is -0.02% for the subset of cryptocurrencies with the highest market capitalization. Similarly, for the group of cryptocurrencies with the lowest trading volume the R^2_{pred} from the `IPCA4` is 0.56% against a 0.05% for the assets with high average trading volume.

Overall, the performance of the IPCA within the context of cryptocurrency markets is broadly similar to the evidence on more traditional asset classes such as equity (see, Kelly et al., 2019). The IPCA offers an especially accurate description of realised returns of "higher-quality" assets. Instead, we see that IPCA produces a higher predictive R^2 for "lower-quality" assets, meaning those cryptocurrencies which are smaller in terms of market capitalization, less liquid and less active from a fundamental blockchain perspective.

A similar pattern emerges for both the PCA and the observable factor models. When we compare the IPCA against both PCA and observable factor models, Table 10 broadly confirms that our dynamic latent factor model provides a more accurate risk-based description of realised and expected returns across different groups of assets. For instance, the R^2_{tot} from the `IPCA4` for the smallest (largest) assets is 9% (26.7%), against a 6.5% (24.3%) from the `IFF7` model. Similarly, for the group of assets with the smaller (larger) trading volume, the `IPCA4` produces a total R^2 of 8.4% (28.7%), against an R^2_{tot} of 5.2% (24.8%) obtained from the `IFF7` model. Consistent with the full-sample results, the PCA represents a rather challenging benchmark when it comes to explain the common variation in the realised returns. However, the IPCA stands out for its predictive performance. For instance, within the group of assets with a lower (higher) number of active addresses, the `IPCA4` produces an R^2_{pred}

of 0.29% (0.13%) against the 0.03% (0.01%) of the `PCA7` and the 0.08% (-0.03%) obtained from the `IFF7`. Similarly, for assets with lower (higher) trading volume, the `IPCA4` produces a predictive R^2 of 0.56% (0.05%) against the 0.22% (0.01%) of the `PCA7` and the 0.21% (-0.12%) generated by the `IFF7`. Overall, the IPCA provides a more accurate risk-based explanation of the variation in expected returns than both PCA and standard observable factor models.

3.4 Weekly returns

Our analysis has thus far focused on the daily return horizon. Given the relatively short history of cryptocurrency markets (see Figure 1), the use of daily returns substantially increases the amount of information that can be used to extract latent factor models and/or to construct observable risk factors. However, daily returns are particularly volatile, especially within the context of cryptocurrency markets. Thus, it is possible that IPCA in part capture noisy fluctuations in the dynamic of individual returns, effects that may be less influential at a lower frequency. In addition, by using weekly returns, while the size of the cross section remains unchanged, the length of the time series is substantially reduced, creating a further challenge for the extraction of latent factors.

As an extension and robustness assessment, we re-analyse the performance of the IPCA model using a weekly aggregation of the returns and individual characteristics. A weekly aggregation of both individual returns and observable risk factors is consistent with some of the existing literature, such as Liu et al. (2022); Cong et al. (2021b). The basic structure for weekly returns is unchanged from the main empirical analysis on daily returns, with the exception that individual returns are aggregated weekly. Given that cryptocurrency markets are operational on a 24/7 basis, the weekly aggregation is defined with a start time of Sunday 00:00:00 UTC. Individual characteristics are also aggregated weekly, where the aggregation depends on the nature of the information. For instance, both the `new add` and `active add` variables are aggregated weekly by summing up the daily observations. The weekly market beta `capm` β is approximated as the average daily value within the week. The same holds for liquidity measures such as `illiq` and `bid-ask` spreads.

Table 11 reports both the in-sample and out-of-sample performance of the restricted IPCA with $\Gamma_\alpha = 0$ versus a static PCA and an instrumented observable factors model.[12] Similarly to the main results, the weekly aggregation seems to favour a small-scale factor model to explain the variation

[12]We choose to report the instrumented observable risk factor model because it represents a more direct comparison with the IPCA. Furthermore, the results from the static version of the observable risk factor model are slightly worse, so that the dynamic version represents a more challenging benchmark for the IPCA.

in both realised and expected returns. The bootstrap test suggests that a two-factor IPCA already provides a risk-based explanation of the returns which is potentially solely based on risk exposures, i.e., $\Gamma_\alpha = 0$ for $K \geq 2$. A more data-driven information criterion suggests that three latent factors capture most of the comovement in the individual returns. Similarly, the relative pricing error on the managed portfolios shows that there is no considerable pricing gain after an IPCA with three latent factors. For these reasons, in the following we consider an `IPCA3` as our baseline specification.

Focusing on the out-of-sample returns, the performance of the IPCA significantly increases. This is potentially due to the higher signal-to-noise ratio of weekly returns compared to daily returns. For instance, the R^2_{tot} (R^2_{pred}) obtained from the `IPCA3` model on weekly individual returns is 19.7% (0.9%), which is double the 11.5% (0.3%) obtained on daily returns. Also the time series and cross sectional R^2 are higher when using weekly returns. For instance, the `IPCA3` fitted on weekly data produces an R^2_{ts} (R^2_{cs}) of 25.2% (14.2%) compared to a more modest 20.3% (8.5%) for the daily returns.

The explanatory power of the static PCA and the instrumented observable factors model also increases when we use weekly returns. The out-of-sample R^2_{tot} from the benchmark `IFF7` model jumps to 15.1% compared to a 8.5% based on daily returns. Similarly, the `PCA7` performance goes from 16.7% for daily returns to 23.2% for weekly returns. Nevertheless, the IPCA performance is still substantially better than both static latent and observable risk factor models at the weekly frequency. This is particularly clear when it comes to explain the variation in the average returns; the `IPCA3` produces an R^2_{pred} of 0.9% against a 0.36% and 0.26% from the `PCA7` and `IFF7`, respectively. The IPCA also provides a much more accurate risk-based representation of the variation in both realised and expected returns of managed portfolios. For instance, the $R^2_{tot,x}$ ($R^2_{pred,x}$) from the `IPCA3` is 56.4% (7.6%) against a 34% (4.5%) and 26.9% (2.3%) obtained from the `PCA7` and `IFF7`, respectively. Perhaps more importantly, the `IPCA3` produces a smaller relative pricing error of 6.6% compared to 46.8% and 68.4% obtained from both competing factor model specifications.

3.4.1 Mean-variance efficiency. Figure 5 reports the average absolute alphas for the conditional IPCA and the competing latent and observable risk factor models. For the instrumented models (`IPCA3` and `IFF7`), the alphas are computed as the time-series average of the period-by-period portfolios residuals. Instead, for the static latent factor model (`PCA7`) and observable factors (`FF7`), the alphas are computed as intercepts from time series regressions of portfolio returns on the factors. All portfolios are re-leveraged to yield 35% weekly volatility.

Similarly to the case of daily returns, the IPCA produces a substantially lower average absolute pricing error, with a 2.9% from the `IPCA3` versus 5.8% and 7.5% obtained from the `PCA7` and `IFF7`, respectively. Furthermore, the estimated alphas from the competing factor models are more clustered around the 45-degree line compared to the IPCA. This suggests that, also at the weekly frequency, the IPCA provides a risk-based explanation of the variation in managed portfolio returns which is more consistent with theoretical underpinnings of mean-variance efficiency.

Delving deeper into the significance of individual managed portfolio alphas, Panel B of Table 7 shows that not only do both `PCA7` and the benchmark `IFF7` have more significant alphas compared to the baseline `IPCA3`, but also that those significant alphas tend to have a much larger value annualized. For instance, the $\alpha_{\texttt{r7\_1}}$ is 6.8% for the `IPCA3`, while is more than 20% across all competing factor models. Similarly, the $\alpha_{\texttt{r30\_1}}$ is 5.6% for the IPCA model versus more than 12% (19%) for the observable factor models (PCA). This indicates that, despite the smaller average absolute pricing error, observable and standard latent factor models still provide a less accurate risk-based representation of cryptocurrency returns.

4 IPCA factors interpretation

Understanding the nature of the IPCA performance is key to a more structural interpretation of the results. In this section, we test for the driving factors in the dynamics of risk exposures and provide an interpretation of the latent factors extracted from the cross section of individual returns based on the IPCA methodology.

4.1 Expected returns and individual characteristics

Assuming the returns dynamics is solely described by individual characteristics, i.e., $\Gamma_\alpha = 0$, the expected returns from the IPCA are defined as $E_t\left[r_{i,t+1}\right] = \widehat{\beta}'_{i,t}\widehat{f}_{t+1}$, with $\widehat{\beta}_{i,t} = \boldsymbol{z}'_{i,t}\widehat{\Gamma}_\beta$ a direct function of $\boldsymbol{z}_{i,t}$, an $L \times 1$ vector of observable cryptocurrency characteristics. As a result, by testing the significance of the l^{th} row in the parameter matrix Γ_β, one can understand the role of each $\boldsymbol{z}_{i,t}$ characteristic for the dynamics of expected returns $E_t\left[r_{i,t+1}\right]$.

We follow Kelly et al. (2019) and implement a bootstrap approach that tests the joint significance for each individual characteristics across K latent factors. Let the l^{th} row in the parameter matrix

$\Gamma_\beta = [\gamma_{\beta,1}, \ldots, \gamma_{\beta,L}]'$ correspond to the loadings on the K factors of the l^{th} characteristic. The null hypothesis is that the entire l^{th} row must be zero; that is, the lth characteristic does not drive the dynamics of the factor loadings. To test this hypothesis, we begin by estimating an unrestricted IPCA model, in which coefficients of Γ_β are not set to zeros, and save the estimated model parameters $\{\widehat{\gamma}_{\beta,l}\}_{l=1}^{L}$, latent factors $\{\widehat{f}_t\}_{t=1}^{T}$, and managed portfolio residuals $\{\widehat{d}_t\}_{t=1}^{T}$. For each characteristic l, we then compute the Wald-type statistic in the form $\widehat{W}_{\beta,l} = \widehat{\gamma}'_{\beta,l}\widehat{\gamma}_{\beta,l}$. Next, we use the residuals to resample the managed portfolio returns under the restriction $\gamma_{\beta,l} = 0_{K\times 1}$.[13] Then, we re-estimate the IPCA model using these synthetic portfolio returns and compute the bootstrap test statistic $\widehat{W}^b_{\beta,l}$ for the b^{th} bootstrap draw. For the l^{th} characteristic, the p-value of the null hypothesis test equals the fraction of bootstrapped $\widehat{W}^b_{\beta,l}$ statistics exceeding the empirical value $\widehat{W}_{\beta,l}$. Because all of the characteristics are cross-sectionally rank standardized, the reported magnitudes are directly comparable across characteristics.

Table 12 show the p-values for each of the 28 characteristics for five different IPCA specifications, with $K = 2, 3, 4, 5, 6$, based on the full sample. In addition, we report the testing results for the baseline `IPCA3` and `IPCA4` when we split the sample in pre and post Covid-19 outbreak. Finally, the table also reports the testing results for the `IPCA2` and `IPCA3` specifications fitted on the weekly aggregated returns.

Focusing first on the full sample, we find that only a handful of characteristics contribute to explain the dynamics of expected returns as indicated by p-values below the conventional 5% threshold. For instance, for the `IPCA3` specification, variables related to liquidity (`illiq`, `bidask`), past performance (`max7`, and `max30`), and volatility (`rvol`, and `ivol`) are statistically significant at conventional thresholds. The nature of the characteristics that drive the loadings for the `IPCA4` model is similar, albeit there are some differences. For instance, two trading frictions variables, such as `std_to` and `std_vol`, are now significant with a p-value below 0.05. Yet, `illiq`, `max7`, `max30` and `ivol`, are all still significant at conventional levels. Interestingly, the higher the number of factors, the more "sparse" is the nature of the loadings. This suggests an interplay between the role of the latent factors and the characteristics in capturing the common variation in the returns.

[13] Starting from the restricted matrix

$$\widehat{\Gamma}^l_\beta = [\widehat{\gamma}_{\beta,1}, \ldots, \widehat{\gamma}_{\beta,l-1}, 0_{K\times 1}, \widehat{\gamma}_{\beta,l+1}, \widehat{\gamma}_{\beta,L}]',$$

the bootstrap portfolio returns are defined as $\widehat{x}^b_t = Z_t\widehat{\Gamma}^l_\beta\widehat{f}_t + \widehat{d}^b_t$, in which $\{\widehat{d}^b_t\}_{t=1}^{T}$ are the residuals for the b^{th} bootstrap draw.

Except for a few nuances, the testing results over the two sub-sample periods confirm the pattern over the full sample. Liquidity, past performance, and volatility seem to play a major role for the dynamics of the factor loadings. For instance, for the `IPCA4` specification, `illiq`, `max7`, `max30`, and `ivol`, are all significant at 5% levels across both sub-samples. A few changes also occur; for instance, over the pre-Covid 19 period, a more parsimoniou three-factor IPCA model implies that few variables related to liquidity, such as `bid-ask` and `turnover` also drive the dynamics of the factor loadings.

The last two columns of Table 12 shows the testing results for the aggregation to weekly returns. The set of parameters Γ_β for the weekly returns tend to be similar to daily returns. For instance, both `illiq` and `rvol` are significant for the `IPCA3` model estimated at either frequencies. With the exception of the market beta (`capm` β), the testing results at the monthly frequency are mostly consistent with the daily returns, both for the full sample and the sub-samples.

One comment is in order, a simple correlation analysis shows that some of the individual characteristics are potentially quite correlated; for instance, `$volume` is quite correlated with `size`, and `bidask` is quite correlated with `illiq`. As a result, rather than discussing the exact characteristic, our aim is to detect the underlying economic forces that drive the factor loadings. Thus, the results suggest that to a large extent the factor loadings, and therefore expected returns, are primarily affected by liquidity, volatility, and past performance.

4.2 IPCA and observable risk factors

We formally tests whether coupling latent and observable risk factors significantly improves the explanatory power of the IPCA model. We estimate an extended IPCA model of the form

$$r_{i,t+1} = \beta'_{i,t} f_{t+1} + \delta'_{i,t} g_{t+1} + \epsilon_{i,t+1}. \tag{11}$$

with the term $\beta'_{i,t} f_{t+1}$ being the same as in the main IPCA specification. The new term is the portion of the return variation described by the $M \times 1$ vector of observable risk factors g_{t+1}. For consistency, the loadings on both observable and latent risk factors are instrumented using the same set of individual asset characteristics, i.e., $\delta_{i,t} = \boldsymbol{z}'_{i,t}\Gamma_\delta$ where Γ_δ is an $L \times M$ mapping from characteristics to loadings. The estimation of Eq.(11) is a simple extension of the original IPCA in Eq.(4)-(5). That is, the model with nested observable risk factors is mapped to the original IPCA by augmenting the factor specification to include g_{t+1}. A detailed description of the estimation procedure appears in Kelly et al. (2019).

Based on Eq.11 the incremental explanatory power of the observable risk factors g_{t+1} can be evaluated in two ways. First, we directly test for the incremental contribution of g_{t+1} by testing for the significance of the corresponding matrix of parameters Γ_δ. Second, one can compare the asset pricing performance of the IPCA with and without additional observable factors, and directly verify the incremental asset pricing performance over a given IPCA specification.

Panel A of Table 13 formally tests whether the inclusion of observable factors improve over IPCA. We report the results for both a joint test on the inclusion of multiple factors, from `FF1` to `FF7`, and for the inclusion of one single factor. The tests, nest the various sets of observable risk factors outlined in Section 2.2 (represented by columns) with different number of latent IPCA factors (represented by rows). Individual hypothesis tests show that the market portfolio (`mkt`), the `size` factor, and realised volatility (`rvol`) carry some significant loading when added to the `IPCA4` specification we used in the main empirical analysis. Differently, when conditioning on five latent factors, only the market portfolio carry some significant additional effect. When we test jointly the additional information content of observable factors, again the hypothesis test points towards a significant additional explanatory power of the market portfolio for most IPCA specifications.

Despite the marginal significance of the market portfolio returns conditional on the IPCA factors, Panel B of Table 13 shows that none of the observable factors offer an economically relevant improvement of the IPCA's total or predictive R^2. For instance, the R^2_{tot} obtained by including seven observable risk factors to an `IPCA4` model is 13.65% against a baseline 13.37% (labelled as `FF0` in Table 13). More importantly, including observable risk factors to explain the variation in expected returns is actually economically slightly detrimental; for instance, the R^2_{pred} obtained by including seven observable risk factors to an `IPCA4` model is 0.29% against a baseline 0.32%.

Turning to the managed portfolios, as we add more latent factors, the marginal contribution of observable factors to explain the common variation in realised and expected returns is negligible, in fact slightly negative. For instance, the $R^2_{tot,x}$ from the `IPCA4` + `FF7` factor model is 51.2% versus 51.9% obtained from the baseline `IPCA4` model. Similarly, an IPCA with four latent factors produces a predictive $R^2_{pred,x}$ of 1.93% versus a 1.77% obtained from an expanded model including seven observable factors. Overall, the evidence shows that adding more observable factors does not materially improve the ability of the IPCA to provide a risk-based explanation of either realised or expected returns. This is consistent with some of the existing results within the context of traditional equity (see, e.g., Kelly et al., 2019), corporate bond (see, e.g., Kelly et al., 2022) and option markets (see, e.g., Büchner and Kelly, 2022).

4.3 Latent factors and asset characteristics

Because the factors in the IPCA framework are not ordered and are only identifiable up to a rotation, creating a detailed interpretation of the individual factors is problematic, perhaps even inappropriate. Moreover, we caution that any labeling of the factors is imperfect, because each factor is influenced to some degree by all of the characteristics, and the orthogonality condition implies that none of the latent factors will match an exact characteristic. Nonetheless, in this section, we build upon the intuition of Ludvigson and Ng (2009) and provide an interpretation of the latent factors based on the marginal R^2_{marg} of a univariate regression of each of the 28 different managed portfolios onto each estimated IPCA factor, one at a time, using the full sample of observations.

For the ease of exposition, in Figure 6 we report the results for four IPCA specifications with $K = 2, 3, 4, 5$ latent factors. We show the cumulative R^2 for each managed portfolio on each factor as a measure of correlation. Focusing on the two latent factors from the `IPCA2` model, the first factor is primarily correlated with the `capm` β and partly with on-chain network activity, although on a smaller magnitude compared to the second factor. The latter more strongly correlates with liquidity, volatility, trading frictions, and the `VaR(5%)`. Overall, one could identify Factor 2 in the `IPCA2` as a market inefficiency factor.

The top-right panel of 6 reports the R^2 of the regressions of individual managed portfolios on the `IPCA3` three latent factors. Again, Factor 1 is primarily correlated with market risk and partly with on-chain activity. The second factor seems to correlate for the most part with short-term reversal `r2_1` and `bm`, in that the marginal R^2 for is higher than the one corresponding to Factor 1 or 3. Indeed, while `max7` is also quite correlated with Factor 2, the lion's share in terms of correlation is played by Factor 3. The same applies for `$volume` and `size`. As a result, Factor 2 within the `IPCA3` specification seems to be primarily related to valuations and short-term performances. Factor 3 shows a much more heterogeneous correlation, in particular in relation to trading friction measures and volatility.

The bottom-left panel of Figure 6 reports the marginal R^2 obtained regressing the managed portfolios and extracted factors from an `IPCA4` model. As we add one more latent factor, the identification becomes slightly more clear-cut. For instance, Factor 1 still primarily correlates with the market beta. However, Factor 2 now is more clearly related to two measures of short-term reversal, such as `max7` and `r2_1`. With the exception of `turnover`, Factor 3 is primarily correlated with measures value, such as `bm`, `size`, and `$volume`, and growth such as `new add` and `active add`. In other words, Factor 3 can be interpreted as a "value" factor as it combines both valuation and growth aspects in the spirit of the Fama-French HML portfolio. Finally, Factor 4 is mostly correlated with both past performance, liquidity (as proxied by `bidask` and `illiq`) and measures of trading frictions, such as `std_to` and `std_vol`.

Finally, the bottom-right panel of Figure 6 reports the R^2 from the auxiliary regressions on the `IPCA5` five latent factors. Factor 1 is primarily correlated with exposure to market risk. By adding a fifth latent factor, the identification of Factor 2 almost uniquely coincides with intermediate momentum `r30_14`. Similarly, Factor 3 is almost unequivocally correlated with both `max7` and `max30`. With the exception of `turnover` and `$volume`, Factor 4 is primarily correlated with measures value, such as `bm`, and `size`, and growth such as `new add` and `active add`. Interestingly, Factor 4 is also "contaminated" by measures of past performances. Nevertheless, past performances are more strongly related to Factor 5. The latter also strongly correlates with measures of volatility, liquidity, and downside risk, as proxied by `VaR (5%)`.

5 Do equity and cryptos share risk factors?

We expand on our analysis of factor models within cryptocurrency markets by asking whether factors from cryptocurrency market price equity, and vice-versa. This question builds on previous literature, including Liu and Tsyvinski (2021) and Bianchi et al. (2022), who emphasize that full market integration should imply that both markets share the same factors and factor premiums. We leverage the flexibility of the IPCA approach and test the significance of the additional information content that equity risk factors brings to explain the common variation in realised and expected cryptocurrency returns.

We begin by formally testing whether coupling observable equity factors with the latent factors extracted from the cross section of cryptocurrency returns significantly improves the explanatory power of the IPCA. We estimate an extended IPCA model as in Eq.(11), with the term $\beta'_{i,t} f_{t+1}$ being the same as in the main IPCA specification. The new term $\delta'_{i,t} g_{t+1}$ represents the portion of the return variation described by the $M \times 1$ vector of equity risk factors g_{t+1}. For consistency, the loadings on the equity risk factors $\delta_{i,t} = \boldsymbol{z}_{i,t}'\Gamma_\delta$ are instrumented using the same set of individual cryptocurrency characteristics $\boldsymbol{z}_{i,t}$, where Γ_δ is an $L \times M$ mapping from characteristics to loadings. This allows us to give a more structural interpretation of the integration between the equity and the cryptocurrency markets to the extent that it actually exists in the data. As a matter of fact, in this setting the betas on cryptocurrency returns on equity factors take the interpretation of *empirical hedge ratios* (see, e.g., Kelly et al., 2022).

We consider as g_{t+1} the Fama and French (2015) five-factor model: the excess return on the market (MKT), the size factor (SMB), value (HML), profitability (RMW) and the investment factor (CMA). We follow Kelly et al. (2019) and construct a test of the incremental explanatory power of equity risk factors after controlling for the baseline IPCA specification. The null hypothesis is $\mathcal{H}_0 : \Gamma_\delta = \mathbf{0}_{L\times M}$ from which we construct a Wald-like test statistic as $W_\delta = \text{vec}\left(\widehat{\Gamma}_\delta\right)'\text{vec}\left(\widehat{\Gamma}_\delta\right)$. W_δ measures the distance between the model with and without the equity risk factors g_{t+1}. If W_δ is large relatively to sampling variation, one can conclude that the equity risk factors carry significant information for the variation of cryptocurrency returns. The sampling variation estimates, and therefore the p-values, are obtained by using the same wild bootstrap method as in Section 4.1.[14] One comment is in order. Unlike equity, cryptocurrencies are traded on a 24/7 basis. This means that there are some discrepancies in the numbers of observations between cryptocurrency and equity returns. We match both samples by indexing to equity dates, that is, for those days for which we do not have equity returns available, we discard the cryptocurrency returns.

Panel A of Table 14 reports the testing results both when using one factor at a time – considering separately the contribution of MKT, SMB, HML, RMW, CMA – and when adding each factor cumulatively – from a one-factor model (F1) to a five-factor model (F5). Irrespective of being added one at a time or all of them together, observable equity factors are redundant as we add IPCA factors. None of the Fama-French factors are statistically significant at conventional levels after controlling for the commonality in realised individual returns as captured by the IPCA latent factors.

Panel B of Table 14 shows that none of the equity risk factors offer an economically significant improvement over the IPCA's total or predictive R^2. For instance, by adding the five Fama-French

[14] First we construct residuals of managed portfolios $\widehat{d}_{t+1} = Z_t'\widehat{\epsilon}^*_{t+1}$ from the estimated model. Then, for each iteration b, we resample the portfolio returns imposing the null hypothesis $\Gamma_\delta = 0$. Next, for each bootstrap sample, we re-estimate Γ_δ and construct the associated test statistic $\tilde{W}^b_\delta$. Finally, we compute the p-value as the fraction of $\tilde{W}^b_\delta$ that exceeds W_δ.

factors to the baseline `IPCA4` model, both the R^2_{tot} and the R^2_{pred} remains virtually unchanged. The same holds when we use managed portfolios as test assets; the $R^2_{tot,x}$ and $R^2_{pred,x}$ obtained from the `IPCA4` essentially do not change by adding equity risk factors. Overall, Table 14 suggests that once we control for IPCA latent factors, the information content of equity risk factors to explain the variation in realised and expected cryptocurrency returns is negligible.

Overall, by leveraging on the flexibility of the IPCA framework and delving deeper into the systematic variation in cryptocurrency returns and the joint factor structure shared by equity and cryptos, we provide evidence that once we condition for the common variation in individual cryptocurrency returns, equity risk factors do not bring economically valuable information on both realised and expected cryptocurrency returns. Our results expand those of some of the previous literature (see, e.g., Liu and Tsyvinski, 2021; Bianchi et al., 2022), both by considering the post Covid-19 period and by studying highly noisy and volatile daily returns. The latter in particular poses a particular challenge for the latent factor model in extracting fundamental pricing information based on cryptocurrency characteristics.

5.1 Factors spanning regressions

The relatively low additional information content of equity risk factors when added to IPCA factors extracted from cryptocurrency returns, does not mean that the two markets are necessarily segmented. Risk factors can be highly correlated and therefore capture similar sources of risk. We calculate a set of correlations between the Fama and French (2015) equity factors and the latent factors extracted from the IPCA on cryptocurrency returns. This provides additional, albeit indirect, evidence on the similarities and differences in the pricing kernel between cryptocurrencies and equity markets. Because latent factors can only be identified up to a rotation, we assess the correlations between crypto and equity factors using a series of spanning regressions, that is, we regress each of the latent cryptocurrency factors individually on all of the Fama-French equity factors.

The first three columns of Table 15 shows the regression results when the dependent variable is the first three latent factors obtained from the baseline `IPCA4` model. Panel A reports the regression results for the full sample of observations. With the exception of Factor 1, none of the five Fama-French factors is statistically significant at the conventional 5% threshold. However, the equity market factor is indeed significantly correlated with Factor 1, with a spanning regression coefficient that is significant at the 1% level. Nevertheless, the constant, meaning the unexplained returns, is strongly significant for all the three latent factors.

The multiple correlation coefficients ($\sqrt{R^2}$) is also quite low, 26% for Factor 1 to 5% for Factor 3. Panel B of Table 15 reports the results for the period from March 2020 to the end of the sample. Although the multiple correlation coefficients increase, the equity market factor remains the only one significantly correlated with Factor 1. Similar to the full sample, neither Factor 2 nor Factor 3 extracted from the `IPCA4` are correlated with conventional Fama-French equity risk factors.

In order to gain a better perspective on the spillover effects between equity and cryptocurrency markets, we now regress the Fama-French equity risk factors on each one of the observable cryptocurrency factors used in the main empirical analysis (see Section 2.2 for a description). The central part of Table 15 reports the results. Three interesting facts emerge; first, with the exception of the `bm` portfolio, the constant, meaning the unexplained returns, is strongly significant for all cryptocurrency factors. Interestingly, although the intercept for the `bm` portfolio is not significant, none of the equity risk factors have significant spanning regression coefficients either. This suggests that `bm` produces average returns which are neither correlated with equity risk factors, nor significantly different from zero.

Second, the cryptocurrency and equity market factors are positively and significantly correlated. This confirms the conventional wisdom that the aggregate market trend in both asset classes may be correlated.[15] Related to that, the third fact that emerges from the spanning regressions is that the correlation between observable equity and cryptocurrency risk factors tend to increase in the second half of the sample. All of the $\sqrt{R^2}$ increase, with the cryptocurrency market factor now significantly correlated with the HML portfolio at a 1% level. Nevertheless, and consistent with the IPCA spanning regressions (first three columns), the correlation between equity and cryptocurrency factors is far from perfect For instance, with the exception of the `mkt` factor, all of the $\sqrt{R^2}$ are below 20%.

Intuitively, the correlation between IPCA and equity risk factors seem rather small. However, to quantify what actually "small" means in this setting, one needs to look at the correlation between the IPCA and the observable cryptocurrency factors. On the one hand, this gives us a benchmark to gauge the correlation between IPCA and equity factors. On the other hand, this allows us to understand how much overlap there is in the information content between IPCA and traditional cryptocurrency

[15] See for instance https://www.bloomberg.com/news/articles/2022-01-25/bitcoin-is-moving-in-tandem-with-stocks-like-never-before-chart.

factors á-la Fama and French. The last three columns of Table 15 report the spanning regression results where we regress the first three latent factors from an `IPCA4` on the seven observable risk factors used in the main empirical analysis.

Although the regression intercepts are still strongly significant, the multiple correlation coefficients are much larger compared to the equity risk factors. For instance, the $\sqrt{R^2}$ of the three latent IPCA factors on all seven observable cryptocurrency portfolios are 82%, 19% and 31%, respectively. Interestingly, the correlation between IPCA and observable cryptocurrency factors is quite stable when we focus on the second half of the sample. Despite a higher correlation though, the unexplained factor returns are still large and significant. This suggests that (1) equity risk factors do not provide useful information about the latent IPCA factors, and (2) observable cryptocurrency factors do indeed provide useful information, although they do not span the latent factor space with sufficient accuracy.

Overall, Table 15 provides some insight on the intersection between equity and cryptocurrency markets. On the one hand, the relatively low correlation between crypto and equity risk factors that permeats from the regression analysis, suggests that market segmentation may still potentially represent an impediment to cross-asset fundamental spillovers between equity and cryptocurrencies, compared to other asset classes such as bond (see, e.g., Kelly et al., 2022), foreign exchange and commodities (Asness et al., 2013). On the other hand, the presence of a moderate correlation between markets, as shown for instance by the first IPCA latent factor and the equity market portfolio, potentially suggests that investors' hopes on the "diversification" benefits of cryptocurrencies may have been ill-posed (see, e.g., Baek and Elbeck, 2015; Yermack, 2015; Biais et al., 2020; Liu and Tsyvinski, 2021).

As a complementary evaluation of the cross-asset pricing performance, in Appendix C we look at the pricing error on equity portfolios based on a Fama-French five-factor model compared to the baseline `IPCA3` crypto factor model. We measure the extent to which the latent factors from the IPCA fits on cryptocurrency returns produce comparable average alphas vs. the Fama-French equity factor model. We consider as test assets 25 portfolios sorted on size and book-to-market as test assets. Figure C1 shows the results. The average absolute pricing error is 3.4% annualized when we use the five-factor Fama and French (2015) model. This is almost a tenfold smaller than when we use the IPCA factors extracted from the cross section of individual cryptocurrency returns: the average absolute alpha from the `IPCA4` is 29.5% annualised. In addition, for the `IPCA4` model, the alphas are clustered around the 45-degree line, which suggests that the factors extracted from the panel of cryptocurrency returns do not provide significant pricing information for equity markets. Results are similar for 25 portfolios sorted by size and momentum (bottom panels).

6 Conclusion

We build upon an instrumented principal component analysis and show that the characteristics/expected return relationship within the context of cryptocurrency markets is driven by compensation for the exposure to latent risk factors. Our approach uses both returns and characteristics to jointly estimate a set of latent factors that better explain the total variation in realised and expected returns. As a result, our model provides a dynamic characterization of expected returns and risk premiums without taking a dogmatic stand *a priori* on (1) which characteristics matter and (2) which test assets should be used to understand the risks and returns in cryptocurrency markets. We see both these properties as crucial within the context of this fast growing, and arguably still relatively unknown asset class.

Empirically, we show that a parsimonious IPCA factor model outperforms a benchmark observable risk factors model built upon prior literature. That is, the IPCA explains a larger fraction of daily realized and expected cryptocurrency returns and yields better predictions that result in smaller pricing errors. These results hold for both individual asset returns and managed portfolios, during both pre and post COVID-19 crisis periods, and for weekly aggregation of returns and characteristics. In addition, by comparing equity and cryptocurrency factors, within a self-contained asset pricing framework, our results highlight an increasing, although not perfect, correlation between equity and cryptocurrency risk factors. Nevertheless, conditional on the IPCA factors extracted from cryptocurrency returns, none of the standard equity risk factors provide significant information to understand risk compensation in cryptocurrency markets.

References

Abadi, J. and Brunnermeier, M. (2018). Blockchain economics. Technical report, National Bureau of Economic Research.

Abdi, F. and Ranaldo, A. (2017). A simple estimation of bid-ask spreads from daily close, high, and low prices. *The Review of Financial Studies*, 30(12):4437–4480.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets*, 5(1):31–56.

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2006). The cross-section of volatility and expected returns. *The journal of finance*, 61(1):259–299.

Asness, C., Moskowitz, T., and Pedersen, L. (2013). Value and momentum everywhere. *The Journal of Finance*, 68(3):929–985.

Babiak, M., Bianchi, D., and Dickerson, A. (2022). Trading volume and liquidity provision in cryptocurrency markets. *Journal of Banking and Finance*, forthcoming.

Baek, C. and Elbeck, M. (2015). Bitcoins as an investment or speculative vehicle? a first look. *Applied Economics Letters*, 22(1):30–34.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bali, T. G., Cakici, N., and Whitelaw, R. F. (2011). Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of financial economics*, 99(2):427–446.

Biais, B., Bisiere, C., Bouvard, M., Casamatta, C., and Menkveld, A. J. (2020). Equilibrium bitcoin pricing. *Available at SSRN 3261063*.

Bianchi, D., Guidolin, M., and Pedio, M. (2022). The dynamics of returns predictability in cryptocurrency markets. *European Journal of Finance*, (forthcoming).

Brauneis, A., Mestel, R., Riordan, R., and Theissen, E. (2021). How to measure the liquidity of cryptocurrency markets? *Journal of Banking & Finance*, 124:106041.

Büchner, M. and Kelly, B. (2022). A factor model for option returns. *Journal of Financial Economics*, 143(3):1140–1161.

Chiu, J. and Koeppl, T. V. (2017). The economics of cryptocurrencies–bitcoin and beyond. *Available at SSRN 3048124*.

Chordia, T., Subrahmanyam, A., and Anshuman, V. R. (2001). Trading activity and expected stock returns. *Journal of financial Economics*, 59(1):3–32.

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of finance*, 66(4):1047–1108.

Cong, L. W. and He, Z. (2019). Blockchain disruption and smart contracts. *The Review of Financial Studies*, 32(5):1754–1797.

Cong, L. W., He, Z., and Li, J. (2021a). Decentralized mining in centralized pools. *The Review of Financial Studies*, 34(3):1191–1235.

Cong, L. W., Karolyi, G. A., Tang, K., and Zhao, W. (2021b). Value premium, network adoption, and factor pricing of crypto assets. *Working Paper*.

Cong, L. W., Li, Y., and Wang, N. (2021c). Tokenomics: Dynamic adoption and valuation. *The Review of Financial Studies*, 34(3):1105–1155.

Corwin, S. A. and Schultz, P. (2012). A simple way to estimate bid-ask spreads from daily high and low prices. *The Journal of Finance*, 67(2):719–760.

Daniel, K. and Titman, S. (1998). Characteristics or covariances. *Journal of Portfolio Management*, 24(4):24–33.

Daniel, K., Titman, S., et al. (2012). Testing factor-model explanations of market anomalies. *Critical Finance Review*, 1(1):103–139.

Datar, V. T., Naik, N. Y., and Radcliffe, R. (1998). Liquidity and stock returns: An alternative test. *Journal of financial markets*, 1(2):203–219.

De Bondt, W. F. and Thaler, R. (1985). Does the stock market overreact? *The Journal of finance*, 40(3):793–805.

Dobrynskaya, V. (2021). Cryptocurrency momentum and reversal. *Available at SSRN 3913263.*

Easley, D. and O'Hara, M. (2004). Information and the cost of capital. *The journal of finance*, 59(4):1553–1583.

Fama, E. F. and 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. and French, K. R. (2012). Size, value, and momentum in international stock returns. *Journal of financial economics*, 105(3):457–472.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of financial economics*, 116(1):1–22.

Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3):607–636.

Foley, S., Karlsen, J. R., and Putnicnš, T. J. (2019). Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies? *The Review of Financial Studies*, 32(5):1798–1853.

Freyberger, J., Neuhierl, A., and Weber, M. (2020). Dissecting characteristics nonparametrically. *The Review of Financial Studies*, 33(5):2326–2377.

Fu, F. (2009). Idiosyncratic risk and the cross-section of expected stock returns. *Journal of financial Economics*, 91(1):24–37.

Garfinkel, J. A. (2009). Measuring investors' opinion divergence. *Journal of Accounting Research*, 47(5):1317–1348.

George, T. J. and Hwang, C.-Y. (2004). The 52-week high and momentum investing. *The Journal of Finance*, 59(5):2145–2176.

Giglio, S. and Xiu, D. (2021). Asset pricing with omitted factors. *Journal of Political Economy*, 129(7):1947–1990.

Griffin, J. M. and Shams, A. (2020). Is bitcoin really untethered? *The Journal of Finance*, 75(4):1913–1964.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273.

Hendershott, T. and Seasholes, M. S. (2007). Market maker inventories and stock prices. *American Economic Review*, 97(2):210–214.

Huij, J. and Verbeek, M. (2009). On the use of multifactor models to evaluate mutual fund performance. *Financial Management*, 38(1):75–102.

Ilin, A. and Raiko, T. (2010). Practical approaches to principal component analysis in the presence of missing values. *The Journal of Machine Learning Research*, 11:1957–2000.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *The Journal of finance*, 45(3):881–898.

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

Jensen, T. I., Kelly, B. T., and Pedersen, L. H. (2022). Is there a replication crisis in finance? *Journal of Finance*, forthcoming.

Kelly, B. T., Palhares, D., and Pruitt, S. (2022). Modeling corporate bond returns. *Journal of Finance*, forthcoming.

Kelly, B. T., Pruitt, S., and Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 134(3):501–524.

Leirvik, T. (2021). Cryptocurrency returns and the volatility of liquidity. *Finance Research Letters*, page 102031.

Lewellen, J. and Nagel, S. (2006). The conditional capm does not explain asset-pricing anomalies. *Journal of financial economics*, 82(2):289–314.

Lewellen, J., Nagel, S., and Shanken, J. (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial economics*, 96(2):175–194.

Li, T., Shin, D., and Wang, B. (2018). Cryptocurrency pump-and-dump schemes. *Available at SSRN*.

Liu, Y. and Tsyvinski, A. (2021). Risks and returns of cryptocurrency. *The Review of Financial Studies*, 34(6):2689–2727.

Liu, Y., Tsyvinski, A., and Wu, X. (2021). Accounting for cryptocurrency value. *Available at SSRN 3951514*.

Liu, Y., Tsyvinski, A., and Wu, X. (2022). Common risk factors in cryptocurrency. *The Journal of Finance*, 77(2):1133–1177.

Llorente, G., Michaely, R., Saar, G., and Wang, J. (2002). Dynamic volume-return relation of individual stocks. *The Review of Financial Studies*, 15(4):1005–1047.

Ludvigson, S. C. and Ng, S. (2009). Macro factors in bond risk premia. *The Review of Financial Studies*, 22(12):5027–5067.

Makarov, I. and Schoar, A. (2020). Trading and arbitrage in cryptocurrency markets. *Journal of Financial Economics*, 135(2):293–319.

Makarov, I. and Schoar, A. (2021). Blockchain analysis of the bitcoin market. Technical report, National Bureau of Economic Research.

Nagel, S. (2012). Evaporating liquidity. *The Review of Financial Studies*, 25(7):2005–2039.

Novy-Marx, R. (2012). Is momentum really momentum? *Journal of Financial Economics*, 103(3):429–453.

Pagnotta, E. and Buraschi, A. (2018). An equilibrium valuation of bitcoin and decentralized network assets. *Available at SSRN 3142022*.

Routledge, B. and Zetlin-Jones, A. (2021). Currency stability using blockchain technology. *Journal of Economic Dynamics and Control*, page 104155.

Schilling, L. and Uhlig, H. (2019). Some simple bitcoin economics. *Journal of Monetary Economics*, 106:16–26.

Sockin, M. and Xiong, W. (2020). A model of cryptocurrencies. Technical report, National Bureau of Economic Research.

Weber, W. E. (2016). A bitcoin standard: Lessons from the gold standard. Technical report, Bank of Canada Staff Working Paper.

Yang, D. and Zhang, Q. (2000). Drift-independent volatility estimation based on high, low, open, and close prices. *The Journal of Business*, 73(3):477–492.

Yermack, D. (2015). Is bitcoin a real currency? an economic appraisal. In *Handbook of digital currency*, pages 31–43. Elsevier.

Table 1: **Asset characteristics by category**

This table lists 28 characteristics used in our empirical analysis. We group them into four categories: on-chain measures, trading frictions, past returns, and other. We follow Freyberger et al. (2020) and Liu et al. (2022) in the classification of characteristics. We report detailed variable definitions in Appendix B. The data are sampled daily from September 1st 2017 to September 1st 2022, where a day is defined with a start time of 00:00:00 UTC. Daily prices and volume are aggregated across more than 80 different centralised exchanges.

	On-chain measures	
(1)	`new add`	New addresses: The number of unique addresses that appeared for the first time in a transaction of the native coin in the network.
(2)	`active add`	Active addresses: The number of unique addresses that were active in the network either as a sender or receiver. Only addresses that were active in successful transactions are counted.
(3)	`bm`	Network-to-market value: The cumulative number of unique addresses over the current available supply times the current USD price.
	Trading frictions	
(4)	`$volume`	Trading volume: The total amount of coins/tokens transferred across wallets within and across centralised exchanges.
(5)	`size`	Market capitalization: The market capitalization is defined as the product of the current available supply times the current USD price. The current available supply is calculated as the current supply minus the coins that have been burned.
(6)	`rvol`	Realised volatility: The daily realised volatility calculated based on OHLC prices following the methodology propose by Yang and Zhang (2000).
(7)	`bidask`	The bid-ask spread: A daily bid-ask spread calculation based on OHLC prices. It represents the average of the Abdi and Ranaldo (2017) and Corwin and Schultz (2012) approximations.
(8)	`illiq`	Illiquidity ratio: The ratio between the absolute value of the cumulative intraday returns and the daily trading volume expressed in $ (see Amihud, 2002).
(9)	`capm` β	Capm beta: The market beta calculated based on a 30-day rolling window. The market portfolio is calculated as the value-weighted average of the asset returns available at each day t.
(10)	`turnover`	Turnover: the last day trading volume (`$volume`) over the current available supply.
(11)	`dto`	De-trended turnover: The ratio of daily volume (`$volume`) to current available supply minus the daily market turnover and de-trend by its 180 trading days (see Garfinkel, 2009).
(12)	`ivol`	Idiosyncratic volatility: The standard deviation of the residuals from the CAPM based on a 30-day rolling window. The market portfolio is the value-weighted average of the asset returns available at each day t.
(13)	`std_to`	The standard deviation of the residuals from a 30-day rolling window regression of daily turnover on a constant as in Chordia et al. (2001).
(14)	`std_vol`	The standard deviation of the residuals from a 30-day rolling window regression of daily trading volume (`$volume`) on a constant as in Chordia et al. (2001).
(15)	`rel_to_high`	Closeness to the 90-day high: the ratio of the coin price in $ at the end of the previous day over the previous 90 day high price.
(16-17)	`max*`	Maximum daily return in the previous 7 or 30 days following Bali et al. (2011)
(18-19)	`vol shock *d`	Volume shock: the log deviation of trading volume from its trend estimated over a rolling period of 30 or 60 days. The log standard deviation computed over the same rolling window is used to standardise the estimates due to cross-sectional imbalances (see Babiak et al., 2022).
	Past returns	
(20)	`r2_1`	Short-term reversal as in Jegadeesh (1990).
(21-24)	`r*_1`	Cumulative return from 7, 14, 22, and 31 days before the return prediction to one day before.
(25)	`r30_14`	We define intermediate momentum as the cumulative returns from 30 days before prediction to 14 days before.
(26)	`r180_60`	We define long-term reversal is the cumulative return from 180 days before the return prediction to 60 days before.
	Other	
(27)	`capm` α	The excess return from a CAPM calculated based on a 30-day rolling window.
(28)	`VaR(5%)`	The historical Value-at-Risk at 5% calculated based on past 90-day returns.

Table 2: **Descriptive statistics for asset characteristics**

This table reports summary statistics for characteristics and return predictors. For each variable, we report the cross-sectional mean, median, standard deviation and relevant percentiles of the distribution of individual time-series averages. The data are sampled daily from September 1st 2017 to September 1st 2022, where a day is defined with a start time of 00:00:00 UTC. Daily prices and volume are aggregated across more than 80 different centralised exchanges.

					Percentiles					
	Obs.	Mean	Median	Std	1	5	25	75	95	99
On-chain measures										
`new add`	609,399	1616.08	6.73	17931.73	1.06	1.61	3.45	17.00	190.73	34786.19
`active add`	609,399	3674.60	24.32	39004.67	2.15	4.59	12.23	65.62	540.78	66690.87
`bm`	590,652	1.13	0.012	21.55	0.001	0.001	0.011	0.021	0.152	0.531
Trading frictions										
`$volume` ($mln)	541,129	3.72	0.06	33.06	0.00	0.00	0.01	0.34	4.24	55.75
`size`	590,652	16.17	16.04	2.18	11.24	12.87	14.67	17.42	19.81	22.56
`rvol` (%)	595,224	15.05	12.12	12.85	5.02	6.55	9.55	17.33	31.95	49.45
`bidask` (%)	609,399	8.28	7.66	4.22	2.87	4.07	5.99	9.73	14.09	19.81
`illiq`	459,840	50.68	3.81	185.80	0.00	0.04	0.67	25.61	236.70	526.31
`capm` β	597,549	0.93	0.98	0.21	0.20	0.45	0.85	1.06	1.17	1.29
`turnover` (%)	530,054	6.56	0.42	77.65	0.00	0.02	0.11	1.28	6.85	110.87
`dto`	353,936	-0.26	0.05	8.58	-25.18	-0.31	-0.01	0.22	1.88	11.13
`ivol` (%)	597,549	9.17	7.94	4.63	2.70	3.95	5.94	11.49	19.18	23.49
`std_to`	531,011	0.09	0.01	0.65	0.00	0.00	0.00	0.01	0.10	2.24
`std_vol`	542,080	1.29	1.23	0.50	0.49	0.63	0.88	1.66	2.05	2.85
`rel_to_high` (%)	574,244	56.37	55.76	8.90	33.84	42.64	50.83	61.57	70.34	80.66
`max7` (%)	607,029	13.10	11.45	5.87	5.48	7.12	9.27	15.52	25.69	33.78
`max30` (%)	597,944	25.32	22.21	11.36	10.00	12.98	17.67	31.11	48.49	65.26
`vol shock 30d`	462,699	-0.07	-0.07	0.08	-0.32	-0.19	-0.11	-0.03	0.03	0.11
`vol shock 60d`	434,558	-0.11	-0.10	0.18	-0.61	-0.32	-0.16	-0.04	0.08	0.23
Past returns										
`r2_1` (%)	609,398	-0.20	-0.23	0.26	-0.67	-0.52	-0.35	-0.07	0.19	0.54
`r7_1` (%)	607,029	-0.78	-0.95	1.71	-3.78	-2.63	-1.73	-0.24	1.56	7.10
`r14_1` (%)	604,659	-1.34	-1.76	3.61	-7.27	-5.15	-3.25	-0.40	3.84	14.02
`r22_1` (%)	601,104	-2.00	-3.06	6.68	-11.91	-8.90	-5.30	-0.49	7.25	27.13
`r31_1` (%)	597,549	-2.44	-4.06	10.04	-17.03	-12.48	-7.18	-0.15	11.79	45.87
`r30_14` (%)	597,549	-1.79	-2.64	5.72	-10.22	-7.70	-4.62	-0.48	6.10	23.32
`r180_60` (%)	538,694	-3.74	-11.78	53.80	-71.02	-51.40	-27.08	7.16	57.46	261.76
Other										
`capm` α (%)	597,549	-0.10	-0.14	0.33	-0.56	-0.44	-0.24	-0.03	0.28	1.51
`VaR(5%)` (%)	574,244	16.70	14.62	7.17	7.27	9.07	11.85	19.75	31.59	38.69

Table 3: **Observable risk factors**

This table reports descriptive statistics including the mean, standard deviation, Sharpe ratio (annualised), Sortino ratio (annualised), and skewness of the daily returns of portfolios used as a proxy of sources of systematic risks. We report detailed description of each long-short strategy in Section 2.2. The data are sampled daily from September 1st 2017 to September 1st 2022, where a day is defined with a start time of 00:00:00 UTC. Daily prices and volume are aggregated across more than 80 different centralised exchanges.

			Liquidity		Volatility		Past returns						On-chain activity		
	`mkt`	`size`	`bidask`	`illiq`	`rvol`	`ivol`	`r2_1`	`r14_1`	`r22_1`	`r31_1`	`max7`	`max30`	`bm`	`new`	`active`
Mean (%)	0.13	-0.63	-0.84	-0.70	-0.87	-1.05	0.04	0.20	0.13	0.04	-1.34	-1.21	0.20	0.09	0.34
Std (%)	4.32	3.09	9.02	6.33	8.11	8.84	7.13	6.63	6.37	6.18	7.82	8.42	4.60	5.24	4.92
SR (annual)	0.57	-3.90	-1.78	-2.10	-2.06	-2.26	0.10	0.57	0.40	0.12	-3.27	-2.76	0.82	0.32	1.31
Sortino	0.83	-6.03	-3.07	-3.69	-3.56	-3.91	0.16	0.90	0.62	0.18	-5.75	-4.75	1.41	0.41	1.77
Skew	-1.18	-0.30	-0.08	1.03	-0.13	-0.30	-0.51	-0.50	-0.39	-0.49	-0.37	-0.55	1.03	-2.97	-1.05

Table 4: **Characteristics, returns, and market betas**

This table reports the estimated coefficients from a series of panel regressions of individual returns (Panel A) and market betas (Panel B) on 28 characteristics used in the main empirical analysis. We report estimates, robust standard errors, and corresponding p-values. A full description of characteristics and returns is provided in Section 2.1. The data are sampled daily from September 1st 2017 to September 1st 2022, where a day is defined with a start time of 00:00:00 UTC. Daily prices and volume are aggregated across more than 80 different centralised exchanges. We label with ***, **, * those coefficients significant at a 1%, 5%, and 10% confidence level.

Characteristic	**Panel A:** Realised returns								**Panel B:** Market β			
	Pooled OLS				Fixed effects				Fixed effects			
	Estimate	Std. Error	$Pr(>\|t\|)$		Estimate	Std. Error	$Pr(>\|t\|)$		Estimate	Std. Error	$Pr(>\|t\|)$	
`new add`	-0.019	0.023	0.397		-0.033	0.025	0.182		-0.003	0.012	0.826	
`active add`	0.007	0.025	0.780		-0.016	0.030	0.592		0.000	0.016	0.995	
`$volume`	-0.024	0.023	0.300		-0.043	0.029	0.145		0.129	0.019	0.000	***
`illiq`	-0.075	0.012	0.000	***	-0.065	0.012	0.000	***	0.028	0.008	0.001	***
`bidask`	-0.035	0.011	0.001	***	-0.025	0.011	0.024	**	0.006	0.006	0.359	
`size`	0.046	0.015	0.002	*	0.022	0.014	0.100		-0.019	0.012	0.111	
`bm`	-0.038	0.019	0.044	**	-0.148	0.044	0.001	***	-0.080	0.036	0.027	**
`turnover`	0.045	0.013	0.001	***	0.093	0.038	0.015	**	-0.108	0.032	0.001	***
`dto`	-0.003	0.022	0.896		0.012	0.027	0.646		-0.056	0.017	0.001	***
`max7`	0.005	0.005	0.374		0.005	0.006	0.395		-0.011	0.005	0.020	**
`max30`	-0.021	0.010	0.043	**	-0.027	0.010	0.009	***	0.059	0.006	0.000	***
`rel_to_high`	-0.024	0.014	0.080	*	-0.025	0.014	0.075	*	0.062	0.010	0.000	***
`vol shock 30d`	-0.036	0.014	0.010	**	-0.027	0.014	0.046		-0.034	0.009	0.000	***
`vol shock 60d`	0.014	0.013	0.277		0.010	0.013	0.446		-0.023	0.007	0.001	***
capm α	-0.032	0.014	0.021		-0.025	0.014	0.070		0.013	0.008	0.108	
capm β	0.046	0.016	0.003	***	0.043	0.016	0.007	***				
`rvol`	-0.018	0.009	0.051	**	-0.015	0.010	0.116		0.000	0.011	0.993	
`ivol`	0.038	0.018	0.034	**	0.048	0.019	0.011	**	-0.091	0.015	0.000	***
`VaR 5%`	0.008	0.016	0.612		0.012	0.015	0.438		0.074	0.011	0.000	***
`r2_1`	-0.028	0.010	0.003	***	-0.029	0.010	0.003	***	0.003	0.001	0.008	***
`r7_1`	0.015	0.011	0.154		0.014	0.010	0.163		-0.017	0.003	0.000	***
`r14_1`	0.048	0.014	0.001	***	0.044	0.014	0.001	***	-0.005	0.003	0.112	
`r22_1`	0.024	0.015	0.111		0.021	0.015	0.154		-0.005	0.003	0.148	
`r31_1`	-0.104	0.026	0.000	***	-0.102	0.026	0.000	***	0.008	0.012	0.491	
`r30_14`	0.079	0.018	0.000	***	0.072	0.018	0.000	***	-0.001	0.005	0.845	
`r180_60`	0.036	0.010	0.000	***	0.023	0.008	0.006	***	0.008	0.006	0.174	
`std_to`	-0.065	0.017	0.000	***	-0.061	0.017	0.000	***	0.020	0.012	0.099	*
`std_vol`	0.000	0.009	0.976		0.006	0.010	0.584		-0.021	0.010	0.031	
$R^2_{adj}(\%)$	0.063				0.077				1.298			
Obs.	594,837				594,837				583,382			

Table 5: **In-sample asset pricing performance**

This table reports a series of asset pricing performance measures outlined in Section 3. Panel A reports the performance of the restricted ($\Gamma_\alpha = 0$) IPCA and standard PCA models with $K = 1, \ldots, 7$ latent factors. In addition, we report the values of Bai and Ng (2002) information criteria for each specification of latent factor models as well as the p-values for the test of $\Gamma_\alpha = 0$ for IPCA based on a wild bootstrap with 10,000 draws. Panel B reports the performance of observable factor models including one through six portfolios. The models with observable factors are estimated using standard static time series regressions (labelled as `FF`l) or using the IPCA methodology where the loadings are dynamic by instrumenting with characteristics (labelled as `IFF`l). The data are sampled daily from September 1st 2017 to September 1st 2022, where a day is defined with a start time of 00:00:00 UTC. Daily prices and volume are aggregated across more than 80 different centralised exchanges.

Panel A: Latent factors

	IPCA							PCA						
	IPCA1	IPCA2	IPCA3	IPCA4	IPCA5	IPCA6	IPCA7	PCA1	PCA2	PCA3	PCA4	PCA5	PCA6	PCA7
R^2_{tot} (%)	11.07	12.01	12.74	13.37	13.98	14.50	15.08	10.95	12.95	14.52	16.06	17.45	18.69	19.82
R^2_{pred} (%)	0.00	0.15	0.31	0.32	0.33	0.33	0.33	0.00	0.02	0.02	0.04	0.06	0.06	0.06
R^2_{ts} (%)	18.61	18.89	19.14	19.39	19.51	19.63	19.69	18.94	19.30	19.69	20.12	20.54	20.92	21.27
R^2_{cs} (%)	8.98	9.85	10.63	11.32	11.96	12.52	13.13	8.59	10.33	11.75	13.14	14.42	15.59	16.63
$R^2_{tot,x}$ (%)	16.76	33.72	43.36	51.91	53.57	56.51	58.38	13.64	14.27	14.44	16.92	17.48	17.81	18.32
$R^2_{pred,x}$ (%)	0.03	0.49	2.01	1.93	1.96	1.95	1.95	0.01	0.10	0.12	0.37	0.41	0.45	0.47
$R^2_{ts,x}$ (%)	8.18	26.42	36.53	45.16	46.91	49.90	51.94	5.75	6.43	6.60	9.33	9.91	10.27	10.82
$R^2_{cs,x}$ (%)	12.34	25.14	34.92	43.25	45.17	48.62	50.45	10.60	10.87	10.95	12.21	12.47	12.62	13.00
RPE	99.48	84.03	6.43	6.66	5.86	4.70	4.91	99.52	95.55	94.67	83.02	81.00	79.43	78.42
IC	-7.47	-7.58	-7.62	-7.67	-7.58	-7.53	-7.46	-6.73	-6.74	-6.74	-6.77	-6.77	-6.78	-6.78
$H_0: \Gamma_\alpha = 0$ (pval)	0.00	0.00	0.00	0.02	0.07	0.98	0.96							

Panel B: Observable factors

	Static loadings							Instrumented loadings						
	FF1	FF2	FF3	FF4	FF5	FF6	FF7	IFF1	IFF2	IFF3	IFF4	IFF5	IFF6	IFF7
R^2_{tot} (%)	9.53	10.26	10.37	10.52	10.67	10.79	10.87	9.29	9.97	10.05	10.08	10.16	10.20	10.23
R^2_{pred} (%)	-0.01	-0.04	-0.03	-0.03	-0.03	-0.02	-0.02	-0.01	-0.01	-0.01	0.00	0.02	0.03	0.05
R^2_{ts} (%)	16.79	17.48	17.64	17.91	18.10	18.22	18.30	15.95	16.51	16.66	16.73	16.85	16.87	16.90
R^2_{cs} (%)	7.25	8.03	8.12	8.29	8.43	8.53	8.60	7.11	7.83	7.92	7.96	8.04	8.08	8.11
$R^2_{tot,x}$ (%)	11.71	16.46	17.77	18.12	18.82	19.21	19.47	13.12	18.29	19.91	20.13	20.89	21.40	21.76
$R^2_{pred,x}$ (%)	-0.05	0.41	0.40	0.43	0.59	0.67	0.79	-0.06	0.49	0.49	0.51	0.68	0.75	0.88
$R^2_{ts,x}$ (%)	4.53	9.18	10.48	10.88	11.67	12.07	12.34	5.97	11.08	12.73	12.96	13.81	14.34	14.72
$R^2_{cs,x}$ (%)	8.46	12.06	12.33	12.50	12.75	12.80	12.90	9.35	12.96	13.56	13.72	14.03	14.35	14.54
RPE	102.15	81.11	81.61	80.22	72.82	69.08	63.53	101.63	79.51	80.10	79.42	72.58	69.80	64.63

Table 6: **Out-of-sample asset pricing performance**

This table reports a series of out-of-sample asset pricing performance measures outlined in Section 3. Panel A reports the performance of the restricted ($\Gamma_\alpha = 0$) IPCA and standard PCA models with $K = 1, \ldots, 7$ latent factors. Panel B reports the performance of observable factor models including one through six portfolios. The models with observable factors are estimated using standard static time series regressions (labelled as `FF`l) or using the IPCA methodology where the loadings are dynamic by instrumenting with characteristics (labelled as `IFF`l). The data are sampled daily from September 1st 2017 to September 1st 2022, where a day is defined with a start time of 00:00:00 UTC. Daily prices and volume are aggregated across more than 80 different centralised exchanges. Recursive forecasts are carried out by expanding the window of observations starting from March 2nd 2020.

Panel A: Latent factors

	IPCA							PCA						
	`IPCA1`	`IPCA2`	`IPCA3`	`IPCA4`	`IPCA5`	`IPCA6`	`IPCA7`	`PCA1`	`PCA2`	`PCA3`	`PCA4`	`PCA5`	`PCA6`	`PCA7`
R^2_{tot} (%)	9.25	10.22	10.91	11.49	12.09	12.55	13.02	9.30	11.11	12.31	13.62	14.69	15.74	16.77
R^2_{pred} (%)	-0.04	0.28	0.30	0.30	0.30	0.30	0.30	-0.04	-0.03	-0.03	0.01	0.01	0.03	0.03
R^2_{ts} (%)	19.92	20.44	20.24	20.34	20.49	20.59	20.52	21.24	21.34	21.55	21.79	22.01	22.59	22.29
R^2_{cs} (%)	6.22	7.18	7.88	8.47	9.07	9.55	10.04	6.28	8.17	9.36	10.69	11.77	12.77	13.79
$R^2_{tot,x}$ (%)	15.95	30.40	44.92	52.75	55.76	58.73	61.16	13.12	13.50	13.39	15.20	16.14	17.40	17.76
$R^2_{pred,x}$ (%)	0.00	2.89	2.29	2.20	2.28	2.25	2.21	-0.02	0.02	0.06	0.33	0.40	0.47	0.49
$R^2_{ts,x}$ (%)	9.27	23.82	38.61	46.23	49.19	52.03	54.39	6.56	6.97	6.84	8.74	9.69	11.05	11.45
$R^2_{cs,x}$ (%)	10.21	21.15	34.13	41.90	44.76	48.04	50.70	8.86	8.91	8.61	9.36	9.75	10.44	11.08
RPE	101.35	15.39	13.09	10.92	10.16	10.78	10.72	101.99	99.23	98.81	85.34	80.34	74.58	73.97
IC	-7.43	-7.50	-7.61	-7.65	-7.59	-7.54	-7.49	-7.39	-7.28	-7.16	-7.06	-6.95	-6.85	-6.74

Panel B: Observable factors

	Static loadings							Instrumented loadings						
	`FF1`	`FF2`	`FF3`	`FF4`	`FF5`	`FF6`	`FF7`	`IFF1`	`IFF2`	`IFF3`	`IFF4`	`IFF5`	`IFF6`	`IFF7`
R^2_{tot} (%)	7.95	8.26	8.24	8.17	8.14	8.07	7.99	7.88	8.35	8.41	8.38	8.45	8.48	8.50
R^2_{pred} (%)	0.00	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.04	0.05	0.05	0.07	0.07	0.09
R^2_{ts} (%)	19.36	19.20	19.13	19.09	19.08	18.97	18.82	19.03	18.95	19.06	18.99	19.08	19.12	19.08
R^2_{cs} (%)	5.06	5.35	5.32	5.26	5.20	5.11	5.02	4.99	5.45	5.49	5.47	5.52	5.55	5.57
$R^2_{tot,x}$ (%)	10.82	17.04	18.18	18.03	18.61	19.04	19.13	12.72	19.66	20.99	20.75	21.54	22.01	22.23
$R^2_{pred,x}$ (%)	-0.04	0.60	0.61	0.67	0.78	0.83	0.95	-0.09	0.71	0.74	0.80	0.95	0.99	1.14
$R^2_{ts,x}$ (%)	5.06	11.30	12.40	12.23	12.88	13.29	13.38	6.96	13.91	15.21	14.95	15.82	16.27	16.49
$R^2_{cs,x}$ (%)	6.93	11.63	11.73	11.56	11.57	12.37	12.31	7.81	12.82	13.00	12.82	12.75	13.61	13.53
RPE	100.88	72.04	71.66	71.55	68.28	66.07	63.32	100.72	66.24	65.59	65.79	61.99	60.20	57.33

Table 7: **Managed portfolio alphas**

This table reports the daily and weekly alphas of managed portfolios based on the restricted ($\Gamma_\alpha = 0$) IPCA model with $K = 4$ latent factors (`IPCA4`), a static latent factor model with seven factors (`PCA7`), and the observable factor model including seven portfolios with static (`FF7`) or instrumented (`IFF7`) betas. For the conditional IPCA and the observable factor model with instruments, the portfolio alphas are obtained as time-series averages of the period-by-period model residuals. For the static observable and latent factor, the portfolio alphas are obtained as intercepts from a time series regression of portfolio returns on the observable factors. Absolute portfolio alphas with t-statistics greater than 3.0 are highlighted in green print. The data are sampled daily from September 1st 2017 to September 1st 2022, where a day is defined with a start time of 00:00:00 UTC. Daily prices and volume are aggregated across more than 80 different centralised exchanges.

	Panel A: Daily returns								**Panel B:** Weekly returns							
	`IPCA4`		`PCA7`		`FF7`		`IFF7`		`IPCA4`		`PCA7`		`FF7`		`IFF7`	
	α (%)	t-stat	α (%)	t-stat	α (%)	t-stat	α (%)	t-stat	α (%)	t-stat	α (%)	t-stat	α (%)	t-stat	α (%)	t-stat
`new add`	-0.08	-0.82	0.36	3.29	0.44	3.91	0.29	2.93	-1.85	-1.15	-0.17	-0.12	6.36	3.87	6.40	4.04
`active add`	-0.10	-1.12	0.39	3.56	0.45	4.14	0.33	3.36	-1.80	-1.93	0.34	0.25	6.93	4.18	6.99	4.69
`$volume`	0.03	0.49	1.11	6.66	1.13	6.98	0.85	6.44	-2.92	-3.14	3.32	2.46	11.50	5.14	10.59	4.58
`illiq`	0.07	0.70	-0.06	-0.45	-0.06	-0.52	-0.14	-1.47	-2.83	-1.40	2.06	1.08	2.79	2.39	2.56	1.90
`bid-ask`	0.31	3.21	0.31	2.58	0.27	2.31	0.11	1.14	-3.59	-1.78	-0.23	-0.13	2.63	1.50	2.41	1.25
`size`	0.08	1.00	1.37	12.34	1.17	10.16	0.89	9.87	-0.91	-0.93	6.56	4.13	11.13	4.64	10.20	5.98
`bm`	0.27	2.39	1.73	14.70	1.56	12.51	1.22	11.54	1.15	0.81	8.02	3.30	12.79	5.28	10.59	5.39
`turnover`	-0.10	-1.30	0.76	4.88	0.89	5.86	0.67	4.78	-2.48	-1.93	3.91	2.16	12.63	3.94	10.57	3.27
`dto`	0.03	0.25	0.38	3.51	0.37	3.91	0.33	3.39	0.13	0.10	2.34	1.08	3.07	1.36	1.36	0.67
`max7`	0.23	5.74	1.93	13.56	1.88	15.01	1.84	16.30	-4.06	-2.30	-4.66	-1.60	1.92	0.61	1.34	0.45
`max30`	-0.02	-0.26	0.38	2.51	0.32	2.34	0.45	4.05	-3.76	-1.08	-1.10	-0.33	7.21	1.73	5.93	1.55
`rel_to_high`	0.28	3.18	0.60	6.14	0.52	5.10	0.41	4.08	1.66	1.56	7.06	2.32	5.18	2.51	4.90	2.44
`vol shock 30d`	0.47	4.02	1.21	9.45	1.19	9.42	1.14	9.54	-0.93	-0.75	0.29	0.14	1.96	0.89	1.71	0.72
`vol shock 60d`	0.39	3.62	1.10	9.63	1.08	9.17	1.04	9.67	-0.11	-0.09	2.85	1.13	5.23	1.99	4.58	1.72
`capm` α	0.22	2.88	0.25	2.05	0.17	1.51	0.15	1.36	-2.72	-1.74	-5.46	-1.43	3.55	1.00	2.25	0.63
`capm` β	0.34	3.61	0.24	2.44	0.25	2.58	0.21	2.27	-0.95	-0.82	2.17	0.71	4.99	2.92	3.08	1.78
`rvol`	0.13	1.65	0.30	1.90	0.28	2.06	0.15	1.26	-2.90	-1.54	2.80	1.03	11.70	3.08	10.10	2.85
`ivol`	0.07	0.91	0.28	1.78	0.28	2.07	0.10	0.91	-3.63	-1.88	1.84	0.56	10.24	2.61	8.80	2.36
`VaR(5%)`	0.21	2.92	0.55	3.96	0.51	4.34	0.30	2.93	-1.97	-1.24	6.38	2.46	12.24	3.67	11.17	3.60
`r2_1`	-0.28	-3.24	0.61	5.88	0.57	5.20	0.60	6.21	6.85	3.68	27.04	9.16	23.40	8.85	21.31	9.69
`r7_1`	0.16	2.35	0.51	4.92	0.38	3.86	0.46	4.96								
`r14_1`	0.02	0.34	0.38	3.65	0.22	2.10	0.32	3.26	5.537	1.94	21.44	10.65	15.31	5.19	14.71	5.75
`r22_1`	0.03	0.43	0.32	2.97	0.24	2.33	0.29	3.06	5.81	3.40	20.01	9.55	14.07	5.42	13.68	5.60
`r31_1`	0.23	3.10	0.27	2.16	0.16	1.37	0.20	1.86	5.59	3.16	19.40	8.06	12.60	4.86	12.36	4.94
`r30_14`	-0.32	-3.33	0.35	2.65	0.43	3.12	0.39	3.11	-1.85	-1.23	-2.15	-0.65	5.63	1.70	4.14	1.15
`r180_60`	0.12	1.08	0.28	2.29	0.37	2.81	0.39	3.35	-0.59	-0.56	0.59	0.23	7.68	2.37	6.64	1.86
`std_to`	-0.17	-1.83	0.38	3.60	0.55	5.52	0.48	5.09	-1.52	-1.58	3.60	2.16	10.98	5.05	9.60	4.84
`std_vol`	0.00	-0.04	0.14	1.14	0.18	1.39	0.04	0.44	-3.09	-1.18	1.73	0.53	6.04	1.75	4.86	1.62
Avg. $\lvert\alpha\rvert$	0.17		0.59		0.57		0.49		2.86		5.83		8.51		7.51	

Table 8: **Factor tangency portfolios**

This table summarizes the out-of-sample performance of the mean-variance tangency portfolios formed from the IPCA or observable factors. We assume a portfolio volatility target of 5% daily, consistent with the historical volatility of long-short portfolios, and rescale the portfolio weights accordingly using only backward looking information. The performance measures are the daily average returns, the Sharpe ratio, and the alpha from both a CAPM regression and a multi-factor model in which the returns on the tangency portfolios are regressed on the excess returns of the market portfolio (α_{CAPM}) or the seven observable factors used in the main empirical analysis (see Section 2.2) (α_{F7}). We use an in-sample period of two years and then recursively expand the estimation window. The full sample is from September 1st 2017 to September 1st 2022.

	IPCA						PCA						Observable factors with instruments					
Factors	Mean(%)	SR	α_{CAPM}	t_{CAPM}	α_{F7}	t_{F7}	Mean(%)	SR	α_{CAPM}	t_{CAPM}	α_{F7}	t_{F7}	Mean(%)	SR	α_{CAPM}	t_{CAPM}	α_{F7}	t_{F7}
1	-0.14	-0.03	0.00	-0.02	0.17	2.28	-0.18	-0.04	-0.02	-0.21	0.18	2.13	0.15	0.03	0.00	-2.72	0.00	-2.91
2	3.24	0.64	3.23	16.03	2.84	12.16	0.13	0.02	0.20	1.23	0.30	1.66	1.62	0.34	1.61	6.81	-0.01	-0.84
3	4.07	0.84	4.06	25.54	3.73	21.68	-0.06	-0.01	-0.06	-0.24	0.00	0.00	1.63	0.35	1.62	7.07	-0.02	-1.59
4	4.06	0.83	4.05	23.74	3.72	20.11	0.85	0.14	0.92	4.43	0.90	4.00	1.53	0.33	1.52	6.94	-0.03	-1.58
5	4.09	0.84	4.09	23.79	3.75	20.06	0.83	0.14	0.89	4.62	0.91	4.15	1.59	0.33	1.58	7.63	-0.01	-0.26
6	4.05	0.84	4.05	23.96	3.73	20.17	0.99	0.17	0.99	4.19	0.97	3.56	1.67	0.36	1.66	8.44	-0.01	-0.23
7	4.01	0.83	4.01	22.76	3.69	19.50	1.09	0.18	1.09	4.62	1.05	3.86	1.73	0.40	1.72	9.69	0.00	-0.11

Table 9: **Asset pricing performance across sub-samples**

This table reports the asset pricing performance of the models estimated on the sub-samples from September 1st 2017 to March 1st 2020 (Panel A) and from March 2nd 2020 to September 1st 2022 (Panel B). The performance measures are outlined in Section 3. Each panel reports the performance of the restricted ($\Gamma_\alpha = 0$) IPCA models with $K = 1, \ldots, 7$ latent factors, the p-values for the test of $\Gamma_\alpha = 0$ based on a wild bootstrap with 10,000 draws and the values of Bai and Ng (2002) information criteria for each specification. Each panel also reports the performance of instrumented observable factor models including one (IFF1) through seven (IFF7) portfolios. The data are sampled daily where a day is defined with a start time of 00:00:00 UTC. Daily prices and volume are aggregated across more than 80 different centralised exchanges.

Panel A: Sample from October 2017 to March 2020

	IPCA							PCA							Observable factors (static)						
	IPCA1	IPCA2	IPCA3	IPCA4	IPCA5	IPCA6	IPCA7	PCA1	PCA2	PCA3	PCA4	PCA5	PCA6	PCA7	IFF1	IFF2	IFF3	IFF4	IFF5	IFF6	IFF7
R^2_{tot} (%)	14.13	14.95	15.73	16.37	17.00	17.59	18.15	13.59	14.43	15.40	16.37	17.66	18.86	19.76	11.63	12.66	12.76	12.88	12.95	12.98	13.03
R^2_{pred} (%)	0.05	0.25	0.29	0.29	0.29	0.30	0.30	0.05	0.05	0.05	0.05	0.06	0.06	0.06	-0.04	-0.10	-0.10	-0.07	-0.05	-0.05	-0.03
R^2_{ts} (%)	21.27	21.64	21.93	22.24	22.42	22.56	22.76	12.98	12.99	13.01	13.02	13.05	13.08	13.09	2.32	3.94	4.19	4.79	4.76	4.71	4.57
R^2_{cs} (%)	11.74	12.63	13.49	14.20	14.88	15.53	16.10	20.63	21.08	21.63	22.06	22.60	23.05	23.49	9.20	10.21	10.34	10.44	10.52	10.55	10.60
$R^2_{tot,x}$ (%)	17.14	24.14	38.17	49.91	53.01	54.53	56.71	14.05	14.33	14.56	17.06	17.27	17.53	18.06	13.13	16.39	18.24	19.08	19.81	20.02	20.50
$R^2_{pred,x}$ (%)	0.08	0.78	1.47	1.47	1.41	1.38	1.40	0.07	0.11	0.13	0.15	0.14	0.16	0.16	-0.04	-0.02	-0.03	0.07	0.20	0.22	0.33
$R^2_{ts,x}$ (%)	6.44	14.10	30.39	42.97	46.40	48.20	50.58	4.57	4.89	5.11	8.00	8.15	8.39	8.99	4.42	7.25	9.23	10.19	11.02	11.26	11.76
$R^2_{cs,x}$ (%)	14.34	20.59	33.15	43.87	46.98	48.91	51.34	12.44	12.54	12.62	13.40	13.56	13.79	14.23	10.79	14.00	15.04	15.28	15.52	15.53	15.89
RPE	97.18	55.14	9.88	8.43	8.68	9.31	9.22	95.41	93.30	92.08	90.41	90.97	90.25	90.23	103.06	101.45	103.07	98.05	90.42	89.46	84.31
IC	-7.51	-7.48	-7.56	-7.65	-7.60	-7.51	-7.44	-6.76	-6.76	-6.76	-6.79	-6.80	-6.80	-6.81							
$H_0 : \Gamma_\alpha = 0$ (pval)	0.002	0.001	0.002	0.035	0.131	0.183	0.231														

Panel B: Sample from March 2020 to March 2022

	IPCA							PCA							Observable factors (static)						
	IPCA1	IPCA2	IPCA3	IPCA4	IPCA5	IPCA6	IPCA7	PCA1	PCA2	PCA3	PCA4	PCA5	PCA6	PCA7	IFF1	IFF2	IFF3	IFF4	IFF5	IFF6	IFF7
R^2_{tot} (%)	9.30	10.37	11.11	11.75	12.35	12.88	13.41	9.73	12.55	14.60	16.65	18.29	19.85	21.27	7.95	8.50	8.59	8.62	8.73	8.79	8.85
R^2_{pred} (%)	0.01	0.24	0.40	0.40	0.43	0.44	0.44	0.02	0.05	0.06	0.11	0.15	0.16	0.17	0.01	0.12	0.13	0.13	0.15	0.17	0.20
R^2_{ts} (%)	20.83	20.82	20.83	20.80	20.92	20.92	20.87	22.77	23.22	23.67	24.19	24.68	25.14	25.54	18.86	19.02	19.18	19.26	19.45	19.49	19.54
R^2_{cs} (%)	6.26	7.29	8.06	8.72	9.31	9.84	10.39	6.49	9.45	11.61	13.65	15.25	16.79	18.23	5.08	5.62	5.68	5.70	5.79	5.85	5.90
$R^2_{tot,x}$ (%)	16.96	38.22	47.09	51.09	56.82	60.81	62.96	13.91	15.04	15.29	18.43	19.30	20.46	20.96	13.59	21.01	22.44	22.79	23.89	24.57	24.90
$R^2_{pred,x}$ (%)	-0.09	1.47	3.29	3.33	3.25	3.23	3.21	-0.10	0.17	0.21	0.90	1.10	1.23	1.29	-0.10	1.55	1.56	1.56	1.79	1.92	2.11
$R^2_{ts,x}$ (%)	10.23	31.70	40.58	44.47	50.24	54.04	56.44	7.29	8.45	8.72	11.97	12.88	14.08	14.62	7.78	15.20	16.63	17.02	18.20	18.88	19.20
$R^2_{cs,x}$ (%)	10.58	26.04	35.51	39.78	45.50	50.08	52.64	8.83	9.31	9.38	11.16	11.54	12.46	12.76	8.01	12.59	12.79	12.88	13.53	14.36	14.44
RPE	100.29	55.01	7.14	6.91	5.21	3.16	3.40	102.70	95.27	94.13	74.79	69.14	65.42	63.71	100.61	56.78	56.67	56.38	50.70	47.48	43.14
IC	-7.44	-7.62	-7.65	-7.61	-7.62	-7.60	-7.54	-6.69	-6.70	-6.71	-6.75	-6.76	-6.77	-6.78							
$H_0 : \Gamma_\alpha = 0$ (pval)	0.000	0.000	0.003	0.036	0.072	0.105	0.182														

Table 10: **Individual assets quality and IPCA performance**

This table reports the asset pricing performance for individual returns grouped according to three different characteristics. Each day, we sort the cross section of individual assets in quartiles based on market capitalization, number of active addresses or the average daily trading volume, then calculate the total and predictive R^2 for each quartile. We report the results for difference specifications of the IPCA, the static PCA and the instrumented observable factors model. The data are sampled daily from September 1st 2017 to September 1st 2022, where a day is defined with a start time of 00:00:00 UTC. Daily prices and volume are aggregated across more than 80 different centralised exchanges.

	R^2_{tot} (%)												R^2_{pred} (%)											
IPCA																								
	Market cap				Active addresses				Volume				Market cap				Active addresses				Volume			
K	Low	Q2	Q3	High	Low	Q2	Q3	High	Low	Q2	Q3	High	Low	Q2	Q3	High	Low	Q2	Q3	High	Low	Q2	Q3	High
1	5.75	10.41	15.98	26.08	8.50	8.58	10.61	20.91	4.85	13.37	22.88	28.07	-0.02	0.01	0.02	0.03	0.00	0.00	0.01	0.01	-0.02	0.01	0.01	0.03
2	7.42	11.08	16.30	26.08	9.60	9.67	11.57	21.26	6.58	13.50	23.06	28.43	0.37	0.00	0.03	-0.13	0.09	0.19	0.20	0.03	0.35	-0.04	-0.02	0.02
3	8.11	12.01	17.14	26.47	10.43	10.42	12.34	21.74	7.40	14.11	23.22	28.62	0.48	0.23	0.29	-0.03	0.29	0.37	0.36	0.11	0.54	0.15	-0.07	0.05
4	9.01	12.57	17.56	26.66	11.20	11.05	13.03	22.09	8.34	14.41	23.43	28.70	0.48	0.25	0.31	-0.02	0.29	0.38	0.38	0.13	0.56	0.17	-0.06	0.05
5	9.92	13.31	17.79	26.58	12.03	11.72	13.63	22.34	9.34	14.60	23.52	28.60	0.49	0.26	0.31	-0.02	0.30	0.39	0.38	0.12	0.57	0.17	-0.06	0.05
6	10.68	13.83	18.13	26.59	12.80	12.26	14.13	22.55	10.09	14.93	23.65	28.60	0.49	0.26	0.31	-0.02	0.30	0.39	0.38	0.12	0.57	0.18	-0.07	0.05
7	11.47	14.42	18.50	26.61	13.60	12.90	14.73	22.70	11.01	15.26	23.69	28.51	0.48	0.26	0.31	-0.03	0.30	0.38	0.38	0.11	0.57	0.18	-0.07	0.04
PCA																								
	Market cap				Active addresses				Volume				Market cap				Active addresses				Volume			
K	Low	Q2	Q3	High	Low	Q2	Q3	High	Low	Q2	Q3	High	Low	Q2	Q3	High	Low	Q2	Q3	High	Low	Q2	Q3	High
1	5.70	10.18	15.89	25.77	8.38	8.52	10.40	20.86	4.87	13.12	22.28	27.68	-0.01	0.01	0.01	0.01	-0.01	0.00	0.00	0.01	-0.01	0.01	0.00	0.01
2	9.15	12.28	15.98	25.81	9.02	11.79	12.70	20.99	7.35	13.71	22.68	27.75	0.08	-0.08	0.02	0.01	0.02	0.01	0.02	0.01	0.08	0.01	-0.02	0.02
3	11.09	15.08	16.14	25.86	9.73	13.96	14.67	21.40	9.47	14.23	22.76	27.85	0.10	-0.10	0.02	0.01	0.03	0.02	0.02	0.01	0.09	0.00	-0.02	0.02
4	13.63	16.76	16.33	25.96	11.52	16.24	15.81	21.79	11.48	14.41	22.99	28.07	0.15	-0.11	0.00	0.02	0.01	0.07	0.03	0.02	0.18	-0.02	-0.05	0.02
5	15.75	17.84	16.58	26.08	14.45	18.08	16.38	21.93	13.28	14.81	23.13	28.16	0.18	-0.12	-0.02	0.02	0.02	0.11	0.04	0.02	0.20	-0.03	-0.05	0.01
6	17.71	19.06	16.84	26.15	15.99	19.84	17.27	22.30	15.01	15.06	23.18	28.24	0.21	-0.14	-0.04	0.02	0.04	0.11	0.05	0.00	0.21	-0.03	-0.06	0.01
7	19.35	20.26	17.15	26.26	17.25	21.45	18.18	22.61	16.56	15.49	23.50	28.35	0.22	-0.15	-0.03	0.02	0.03	0.11	0.05	0.01	0.22	-0.03	-0.05	0.01
IFF																								
	Market cap				Active addresses				Volume				Market cap				Active addresses				Volume			
K	Low	Q2	Q3	High	Low	Q2	Q3	High	Low	Q2	Q3	High	Low	Q2	Q3	High	Low	Q2	Q3	High	Low	Q2	Q3	High
1	5.07	8.43	12.87	22.44	7.46	7.29	8.63	17.48	4.13	11.03	18.39	23.59	0.04	-0.04	-0.06	-0.08	0.00	0.00	-0.02	-0.05	0.03	-0.04	-0.04	-0.10
2	6.45	8.52	13.16	22.77	8.24	8.05	9.31	17.87	5.08	11.68	19.14	24.04	0.20	-0.15	-0.17	-0.21	0.02	0.04	-0.03	-0.13	0.16	-0.14	-0.16	-0.26
3	6.45	8.56	13.33	23.14	8.25	8.10	9.39	18.12	5.11	11.73	19.33	24.33	0.20	-0.14	-0.16	-0.20	0.03	0.04	-0.03	-0.12	0.17	-0.13	-0.15	-0.24
4	6.45	8.58	13.35	23.36	8.26	8.12	9.41	18.23	5.13	11.76	19.36	24.47	0.20	-0.14	-0.15	-0.18	0.03	0.05	-0.02	-0.11	0.17	-0.13	-0.14	-0.23
5	6.44	8.60	13.41	23.92	8.25	8.14	9.49	18.52	5.16	11.82	19.48	24.72	0.20	-0.12	-0.11	-0.12	0.04	0.06	0.00	-0.07	0.18	-0.10	-0.11	-0.17
6	6.46	8.61	13.46	24.13	8.26	8.18	9.53	18.61	5.20	11.87	19.52	24.75	0.21	-0.11	-0.09	-0.10	0.05	0.07	0.01	-0.05	0.19	-0.09	-0.10	-0.16
7	6.46	8.62	13.51	24.34	8.28	8.20	9.55	18.72	5.24	11.90	19.52	24.84	0.22	-0.09	-0.06	-0.07	0.08	0.09	0.04	-0.03	0.21	-0.06	-0.08	-0.12

Table 11: **Asset pricing performance on weekly returns**

This table reports the asset pricing performance of the models estimated on weekly aggregate returns and characteristics. The performance measures are outlined in Section 3. We report the performance of the restricted ($\Gamma_\alpha = 0$) IPCA models with $K = 1, \ldots, 7$ latent factors, the static PCA with $K = 1, \ldots, 7$, and the instrumented observable factors model with up to seven factors as outlined in Section 2.2. Panel A reports the in-sample results with the p-values for the test of $\Gamma_\alpha = 0$ based on a wild bootstrap with 10,000 draws and the values of Bai and Ng (2002) information criteria for each specification. Panel B reports the out-of-sample performances where recursive forecasts are carried out by expanding the window of observations starting from March 2nd 2020. The data are sampled weekly from September 1st 2017 to September 1st 2022, where a week is defined with a start time of Sunday 00:00:00 UTC. Prices and volume are aggregated across more than 80 different centralised exchanges.

Panel A: In-sample

	IPCA							PCA							Instrumented observables						
	IPCA1	IPCA2	IPCA3	IPCA4	IPCA5	IPCA6	IPCA7	PCA1	PCA2	PCA3	PCA4	PCA5	PCA6	PCA7	IFF1	IFF2	IFF3	IFF4	IFF5	IFF6	IFF7
R^2_{tot} (%)	20.13	22.08	22.98	23.74	24.43	25.00	25.53	20.27	23.80	26.07	27.70	29.70	31.52	33.14	15.81	18.53	18.67	18.77	19.04	19.20	19.32
R^2_{pred} (%)	0.00	0.95	1.05	1.05	1.06	1.05	1.05	0.02	0.83	0.88	0.89	0.91	0.92	0.93	-0.12	-0.24	-0.20	-0.19	0.01	0.01	0.05
R^2_{ts} (%)	26.86	27.13	27.60	27.95	28.51	28.90	29.31	27.99	29.10	29.90	31.51	32.14	32.97	33.66	21.77	24.16	24.32	24.45	24.67	24.73	24.81
R^2_{cs} (%)	14.73	16.49	17.51	18.25	18.96	19.53	20.11	14.40	17.18	19.28	20.55	22.48	23.96	25.36	9.77	12.69	12.87	12.99	13.33	13.46	13.56
$R^2_{tot,x}$ (%)	23.24	43.16	59.63	62.15	66.52	70.10	73.40	20.88	31.91	32.84	34.50	34.84	35.40	36.10	17.36	27.39	29.56	30.38	32.20	33.45	34.49
$R^2_{pred,x}$ (%)	0.03	3.20	4.29	3.98	4.17	3.89	3.81	0.08	4.45	4.72	4.93	5.09	5.20	5.23	-0.36	1.43	1.59	1.64	2.28	2.32	2.56
$R^2_{ts,x}$ (%)	9.27	30.91	48.94	51.79	56.99	61.52	65.50	7.81	19.80	20.84	22.10	22.47	23.16	23.97	6.64	15.45	17.92	18.96	21.00	22.38	23.58
$R^2_{cs,x}$ (%)	17.02	33.28	49.30	51.96	55.79	59.73	62.99	15.47	23.28	23.94	24.17	24.22	24.71	24.96	9.36	12.96	14.83	15.63	17.27	17.54	18.12
RPE	99.80	22.23	1.20	1.19	0.69	0.78	0.80	98.95	39.50	35.83	33.00	30.75	29.32	28.88	106.40	85.22	83.45	84.28	74.43	75.17	71.05
IC	-5.86	-6.03	-6.24	-6.18	-6.17	-6.16	-6.15	-5.06	-5.21	-5.22	-5.25	-5.25	-5.26	-5.27							
$H_0 : \Gamma_\alpha = 0$ (pval)	0.00	0.18	0.24	0.07	0.06	0.02	0.10														

Panel B: Out-of-sample

	IPCA							PCA							Instrumented observables						
	IPCA1	IPCA2	IPCA3	IPCA4	IPCA5	IPCA6	IPCA7	PCA1	PCA2	PCA3	PCA4	PCA5	PCA6	PCA7	IFF1	IFF2	IFF3	IFF4	IFF5	IFF6	IFF7
R^2_{tot} (%)	16.66	18.94	19.71	20.30	20.98	21.43	21.95	16.42	17.92	18.99	20.46	21.66	22.52	23.26	13.07	14.94	14.99	14.97	15.11	15.17	15.19
R^2_{pred} (%)	-0.35	1.04	0.89	0.90	0.91	0.91	0.93	-0.45	0.08	0.23	0.28	0.33	0.37	0.36	0.11	0.08	0.11	0.10	0.22	0.22	0.26
R^2_{ts} (%)	25.14	25.12	25.12	24.80	25.21	26.00	26.09	27.07	27.77	27.05	27.56	27.02	27.77	27.55	21.57	22.49	22.43	22.36	22.58	22.55	22.79
R^2_{cs} (%)	10.96	13.42	14.16	14.77	15.40	15.82	16.36	10.79	12.27	13.39	14.99	16.27	17.09	17.84	7.57	9.31	9.34	9.32	9.47	9.46	9.49
$R^2_{tot,x}$ (%)	20.98	40.84	56.42	59.42	62.61	64.82	69.06	18.12	28.22	30.23	30.82	31.87	33.97	34.14	15.49	24.76	26.26	25.59	26.62	26.69	26.91
$R^2_{pred,x}$ (%)	-0.21	11.21	7.61	7.31	7.37	7.27	7.58	-0.21	2.81	3.74	4.06	4.25	4.46	4.52	-0.41	1.12	1.43	1.39	1.85	1.88	2.30
$R^2_{ts,x}$ (%)	10.85	30.29	46.79	50.09	53.45	56.07	60.91	7.71	17.50	19.54	20.15	21.24	23.50	23.72	7.46	15.41	16.93	16.38	17.46	17.30	17.49
$R^2_{cs,x}$ (%)	13.76	33.11	46.28	49.94	53.53	55.78	59.80	12.09	18.79	20.74	21.92	22.97	24.68	25.25	8.54	14.82	15.35	15.31	16.58	15.42	15.45
RPE	102.93	16.81	6.60	7.56	6.65	7.05	4.56	103.28	71.40	60.65	55.02	49.42	45.32	46.81	102.53	74.40	71.83	71.78	66.80	70.58	68.37
IC	-5.74	-5.90	-6.07	-6.00	-5.95	-5.87	-5.86	-5.71	-5.70	-5.60	-5.47	-5.35	-5.24	-5.11							

Table 12: **Testing the significance individual characteristics**

This table reports p-values from the bootstrap test for the significance of individual characteristics in the restricted ($\Gamma_\alpha = 0$) IPCA specifications. The `IPCA`l label identifies an IPCA model with l latent factors. We report the results for the full sample – from September 1st 2017 to September 1st 2022 –, the sub-sample separation before and after March 2nd 2020, and the weekly aggregation of the returns. A full description of characteristics and returns is provided in Section 2.1. The significance of the characteristics is color-coded from red (non-significant) to green (significant at conventional levels).

	Full sample					1st sub-sample		2nd sub-sample		Weekly	
	IPCA2	IPCA3	IPCA4	IPCA5	IPCA6	IPCA3	IPCA4	IPCA3	IPCA4	IPCA2	IPCA3
new add	0.387	0.195	0.399	0.651	0.788	0.152	0.338	0.374	0.403	0.306	0.575
active add	0.236	0.277	0.526	0.687	0.710	0.551	0.636	0.641	0.792	0.316	0.527
$volume	0.105	0.091	0.081	0.161	0.382	0.096	0.002	0.071	0.119	0.048	0.041
illiq	0.000	0.000	0.000	0.000	0.000	0.014	0.002	0.000	0.000	0.042	0.040
bidask	0.059	0.039	0.092	0.202	0.372	0.009	0.061	0.102	0.067	0.034	0.075
size	0.057	0.041	0.117	0.244	0.471	0.026	0.101	0.081	0.116	0.848	0.887
bm	0.172	0.092	0.305	0.446	0.582	0.033	0.156	0.115	0.111	0.444	0.666
turnover	0.275	0.149	0.241	0.271	0.552	0.030	0.017	0.265	0.202	0.494	0.765
dto	0.473	0.641	0.741	0.796	0.947	0.746	0.921	0.533	0.225	0.412	0.774
max7	0.458	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.151	0.362
max30	0.425	0.001	0.001	0.005	0.007	0.009	0.004	0.012	0.009	0.310	0.652
rel_to_high	0.301	0.241	0.556	0.775	0.932	0.642	0.502	0.682	0.424	0.476	0.792
vol shock 30d	0.255	0.251	0.501	0.705	0.881	0.839	0.505	0.161	0.164	0.078	0.171
vol shock 60d	0.592	0.621	0.192	0.314	0.531	0.229	0.684	0.221	0.188	0.166	0.391
capm α	0.801	0.866	0.985	0.987	0.895	0.911	0.975	0.661	0.964	0.441	0.438
capm β	0.448	0.919	0.991	0.976	0.972	0.958	0.998	0.683	0.954	0.036	0.009
rvol	0.047	0.016	0.074	0.167	0.251	0.271	0.377	0.023	0.019	0.004	0.000
ivol	0.181	0.041	0.028	0.102	0.073	0.072	0.041	0.021	0.028	0.426	0.143
VaR(5%)	0.331	0.189	0.483	0.702	0.855	0.645	0.949	0.124	0.074	0.156	0.348
r2_1	0.281	0.088	0.288	0.605	0.718	0.042	0.010	0.342	0.301		
r7_1	0.231	0.197	0.012	0.048	0.081	0.087	0.043	0.179	0.113	0.099	0.032
r14_1	0.247	0.187	0.419	0.422	0.304	0.505	0.711	0.255	0.476	0.639	0.009
r22_1	0.321	0.115	0.173	0.441	0.292	0.553	0.243	0.127	0.124	0.510	0.437
r31_1	0.523	0.921	0.986	0.001	0.091	0.905	0.987	0.864	0.000	0.042	0.035
r30_14	0.143	0.089	0.213	0.000	0.000	0.176	0.603	0.115	0.000	0.389	0.262
r180_60	0.179	0.097	0.186	0.316	0.478	0.212	0.579	0.022	0.003	0.091	0.194
std_to	0.232	0.109	0.026	0.002	0.001	0.505	0.821	0.051	0.008	0.712	0.889
std_vol	0.249	0.176	0.029	0.049	0.116	0.375	0.317	0.024	0.029	0.182	0.285

Table 13: **IPCA and observable cryptocurrency factors**

Panel A of the table reports the test results for the null hypothesis that $\Gamma_\delta = 0$ on the instrumented loadings of the observable risk factors (see Eq.11). We report the test results for both a single equity factor at a time, and jointly from one (FF1) to seven (FF7) observable risk factors. Panel B reports the total and predictive R^2 obtained when adding the observable cryptocurrency factors to different IPCA specifications. The full sample is from September 1st 2017 to September 1st 2022.

Panel A: Testing the significance of additional observable factors

	Individual testing							Joint testing						
	mkt	size	r22_1	bm	bidask	rvol	max7	FF1	FF2	FF3	FF4	FF5	FF6	FF7
IPCA1	0.14	0.20	0.32	0.31	0.33	0.06	0.02	0.06	0.00	0.08	0.09	0.00	0.00	0.00
IPCA2	0.56	0.58	0.65	0.57	0.37	0.49	0.02	0.45	0.02	0.09	0.28	0.00	0.00	0.01
IPCA3	0.14	0.11	0.41	0.23	0.21	0.17	0.51	0.04	0.04	0.11	0.22	0.01	0.00	0.78
IPCA4	0.02	0.00	0.19	0.11	0.13	0.04	0.99	0.00	0.14	0.08	0.24	0.11	0.00	0.77
IPCA5	0.02	0.21	0.67	0.07	0.08	0.06	0.98	0.00	0.97	0.32	0.06	0.11	0.00	0.97

Panel B: IPCA explanatory power with the inclusion of additional factors

	R^2_{tot}								R^2_{pred}							
	FF0	FF1	FF2	FF3	FF4	FF5	FF6	FF7	FF0	FF1	FF2	FF3	FF4	FF5	FF6	FF7
IPCA1	11.07	11.21	11.54	11.58	11.60	11.63	11.65	11.68	0.00	0.02	0.17	0.17	0.17	0.17	0.17	0.18
IPCA2	12.01	12.15	12.29	12.32	12.34	12.36	12.38	12.40	0.15	0.17	0.18	0.18	0.18	0.17	0.17	0.18
IPCA3	12.74	12.84	12.95	12.97	12.99	13.02	13.03	13.05	0.31	0.31	0.27	0.28	0.28	0.28	0.28	0.26
IPCA4	13.37	13.46	13.56	13.58	13.60	13.62	13.64	13.65	0.32	0.33	0.31	0.31	0.31	0.31	0.31	0.29
IPCA5	13.98	14.07	14.15	14.17	14.18	14.21	14.22	14.23	0.33	0.33	0.32	0.32	0.32	0.32	0.32	0.32
	$R^2_{tot,x}$								$R^2_{pred,x}$							
	FF0	FF1	FF2	FF3	FF4	FF5	FF6	FF7	FF0	FF1	FF2	FF3	FF4	FF5	FF6	FF7
IPCA1	16.76	20.06	26.50	27.76	27.88	28.21	28.52	28.88	0.03	0.16	1.53	1.55	1.54	1.55	1.55	1.61
IPCA2	33.72	37.02	41.62	42.08	42.19	42.30	42.24	42.16	0.49	0.76	1.64	1.64	1.64	1.63	1.63	1.65
IPCA3	43.36	46.09	44.81	45.09	45.19	45.41	45.38	45.73	2.01	2.00	1.94	1.94	1.93	1.93	1.93	1.93
IPCA4	51.91	52.56	50.55	50.59	50.68	50.89	50.91	51.21	1.93	1.93	1.79	1.77	1.77	1.77	1.76	1.77
IPCA5	53.57	54.62	53.27	53.15	53.19	53.37	53.39	53.47	1.96	1.95	1.90	1.88	1.88	1.88	1.88	1.87

Table 14: **The additional information content of equity risk factors**

Panel A of the table reports the test results for the null hypothesis that $\Gamma_\delta = 0$ on the instrumented loadings (see Eq.11) of the equity risk factors from Fama and French (2012). We report the test results for both a single equity factor at a time, and jointly from one (`F1`) to five (`F5`) equity risk factors. Panel B reports the total and predictive R^2 obtained when adding the equity factors to different IPCA specifications. The full sample is from September 1st 2017 to September 1st 2022.

Panel A: Testing the significance of additional observable factors

	Individual testing					Joint testing				
	MKT	SMB	HML	RMW	CMA	F1	F2	F3	F4	F5
`IPCA1`	0.530	0.480	0.630	0.910	0.690	0.640	0.270	0.720	0.950	0.760
`IPCA2`	0.430	0.510	0.410	0.920	0.540	0.570	0.350	0.610	0.970	0.820
`IPCA3`	0.280	0.500	0.280	0.830	0.570	0.450	0.370	0.590	0.950	0.820
`IPCA4`	0.260	0.390	0.210	0.760	0.510	0.410	0.270	0.350	0.840	0.660
`IPCA5`	0.290	0.340	0.240	0.680	0.480	0.560	0.250	0.320	0.780	0.490

Panel B: IPCA explanatory power with additional factors

	R^2_{tot}						R^2_{pred}					
	F0	F1	F2	F3	F4	F5	F0	F1	F2	F3	F4	F5
`IPCA1`	12.621	12.601	12.611	12.622	12.630	12.641	0.001	0.013	0.013	0.013	0.013	0.015
`IPCA2`	13.546	13.527	13.536	13.547	13.555	13.564	0.151	0.142	0.142	0.142	0.143	0.142
`IPCA3`	14.273	14.255	14.264	14.274	14.282	14.290	0.313	0.331	0.331	0.331	0.332	0.332
`IPCA4`	14.917	14.899	14.907	14.918	14.925	14.934	0.324	0.338	0.338	0.338	0.338	0.338
`IPCA5`	15.519	15.502	15.510	15.520	15.527	15.537	0.327	0.339	0.339	0.339	0.339	0.339

	R^2_{tot}						R^2_{pred}					
	F0	F1	F2	F3	F4	F5	F0	F1	F2	F3	F4	F5
`IPCA1`	18.989	18.837	18.921	19.014	19.072	19.101	0.027	0.085	0.085	0.084	0.077	0.082
`IPCA2`	36.790	36.665	36.736	36.795	36.839	36.916	0.493	0.306	0.310	0.313	0.307	0.310
`IPCA3`	44.762	44.652	44.750	44.748	44.814	44.846	2.011	2.127	2.132	2.129	2.130	2.130
`IPCA4`	53.638	53.644	53.681	53.643	53.614	53.610	1.929	2.027	2.031	2.028	2.028	2.027
`IPCA5`	55.429	55.388	55.442	55.449	55.435	55.431	1.959	2.031	2.034	2.031	2.030	2.031

Table 15: **Factor spanning regressions**

This table reports the results of a number of time-series regressions in which we regress each latent factor from a restricted IPCA model with $K = 3$ on either the five equity factors of Fama and French (2015) or the seven cryptocurrency factors used in the main empirical analysis (see Section 2.2). In addition, we also report the results for a series of time-series regressions in which the dependent variable are the seven observable cryptocurrency risk factors and the independent variables are the Fama and French (2015) equity factors. We report the estimates and label with ***, **, * those coefficients significant at a 1%, 5%, and 10% confidence level, based on robust standard errors. The multiple correlation coefficients in the last row are measured as the square root of R^2. Panel A reports the results for the full sample – September 1st 2017 to September 1st 2022). Panel B reports the results for the second sub-sample from March 2nd 2020 to September 1st 2022.

Panel A: Full sample

Equity vs IPCA/Observable crypto factors

	Dep: IPCA			Dep: Crypto factors						
	Factor 1	Factor 2	Factor 3	`mkt`	`size`	`r22_1`	`bm`	`bid-ask`	`rvol`	`max7`
$\alpha(\%)$	-0.18 **	0.53 ***	0.14 ***	0.17 *	-0.53 ***	0.24 ***	0.05	-1.04 ***	-0.93 ***	-1.53 ***
MKT	-0.70 ***	0.00	0.00	0.67 ***	0.02	-0.20 *	-0.02	0.27	0.06	0.11
SMB	-0.05	0.01	-0.05	0.24	0.08	0.20	-0.08	-0.16	-0.15	0.43
HML	0.18	0.01	0.01	-0.32 **	-0.14	0.13	-0.03	-0.35	-0.37	-0.53 *
RMW	0.43 *	-0.02	-0.06	-0.38 *	0.20	-0.29	0.00	0.98 **	-0.10	0.60
CMF	-0.14	0.01	0.01	0.09	-0.26	-0.30	0.07	1.10 *	0.80	0.59
$\sqrt{R^2}$	0.26	0.03	0.05	0.27	0.08	0.07	0.04	0.08	0.05	0.06

IPCA vs Observable crypto factors

	Dep: IPCA		
	Factor 1	Factor 2	Factor 3
$\alpha(\%)$	0.26 ***	0.51 ***	0.00 ***
`mkt`	-0.37 ***	0.02 ***	0.00
`size`	0.30 ***	0.00	-0.07 ***
`r22_2`	0.04 ***	0.01 ***	-0.01 ***
`bm`	0.02 *	0.00	0.01 **
`bid-ask`	-0.02 *	0.00	0.00
`rvol`	-0.02 ***	0.00	0.00
`max7`	0.01	-0.01 ***	0.00
$\sqrt{R^2}$	0.82	0.19	0.31

Panel B: Sample from March 2nd 2020

Equity vs IPCA/Observable crypto factors

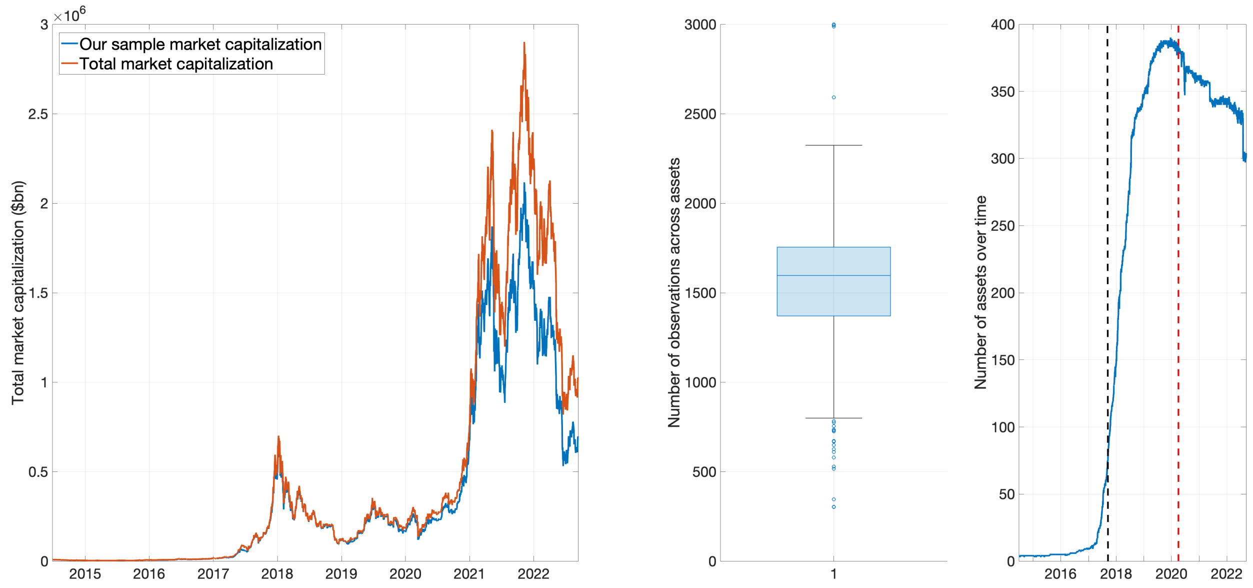
	Dep: IPCA			Dep: Crypto factors						
	Factor 1	Factor 2	Factor 3	`mkt`	`size`	`r22_1`	`bm`	`bid-ask`	`rvol`	`max7`
$\alpha(\%)$	-0.31 **	0.65 ***	0.20	0.20 *	-0.91 ***	0.12 **	-0.126	-1.19 ***	-1.46 ***	-1.60 ***
MKT	-0.86 ***	-0.01	0.00	0.86 ***	0.10	-0.24 **	0.000	0.24	0.13	0.22
SMB	-0.17	0.06	-0.07	0.37 *	0.04	0.33	-0.257 *	0.35	-0.26	0.26
HML	0.30 *	-0.01	0.04	-0.44 ***	-0.12	0.12	0.064	-0.78 ***	-0.49 *	-0.39
RMW	0.26	0.04	-0.08	-0.23	0.17	-0.06	-0.273	1.35 ***	0.12	0.46
CMF	-0.14	0.00	-0.03	0.05	-0.26	-0.51	0.199	1.69 ***	1.39 ***	0.74
$\sqrt{R^2}$	0.41	0.06	0.08	0.41	0.1	0.11	0.09	0.18	0.15	0.09

IPCA vs Observable crypto factors

	Dep: IPCA		
	Factor 1	Factor 2	Factor 3
$\alpha(\%)$	0.09 **	0.63 ***	0.12 ***
`mkt`	-0.91 ***	0.02 **	0.01
`size`	0.23 ***	0.01	-0.08 ***
`r22_2`	0.05 ***	0.01	-0.01
`bm`	0.01	-0.02 ***	0.00
`bid-ask`	0.00	-0.01	0.00
`rvol`	-0.03 ***	0.00	-0.01 *
`max7`	-0.03 *	-0.02 ***	0.00
$\sqrt{R^2}$	0.84	0.25	0.34

Figure 1: **A first look at the data**

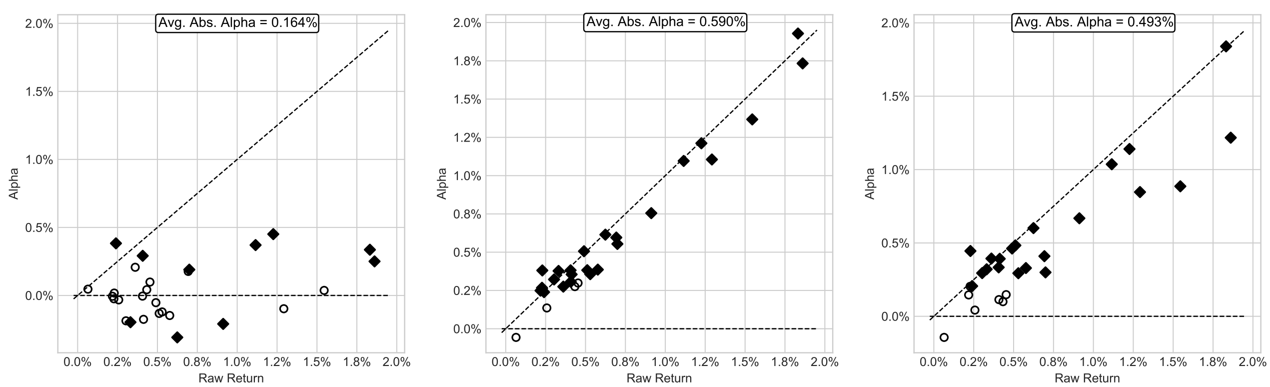
This figure provides a snapshot of the sample used in the main empirical analysis. The left panel compare the market capitalization of cryptocurrencies in our sample and the total market capitalization. The right panels illustrate the time-series and cross-sectional dimensions of the panel of the returns.



(a) Coverage of total market cap (b) Sample composition

Figure 2: **Alphas of managed portfolios**

This figure shows the alphas of managed portfolios based on the set of factors from the restricted ($\Gamma_\alpha = 0$) IPCA model with $K = 4$ factors, a static principal component analysis with $K = 7$ (`PCA7`), and an instrumented factor model with $K = 7$ observable factor portfolios (`IFF7`). For the `IPCA4` and `IFF7` models the alphas are computed as the time-series average of the period-by-period portfolios residuals. For the static `PCA7` the alphas are computed as intercepts from time series regressions of portfolio returns on the latent or observable factors. All portfolios are re-leveraged to yield 5% daily volatility, consistent with the historical volatility of long-short portfolios. Significant alphas with absolute values of t-statistics greater than 2.0 are depicted with filled diamonds, while insignificant alphas are denoted with unfilled circles. All reported values are daily and expressed in percentage.

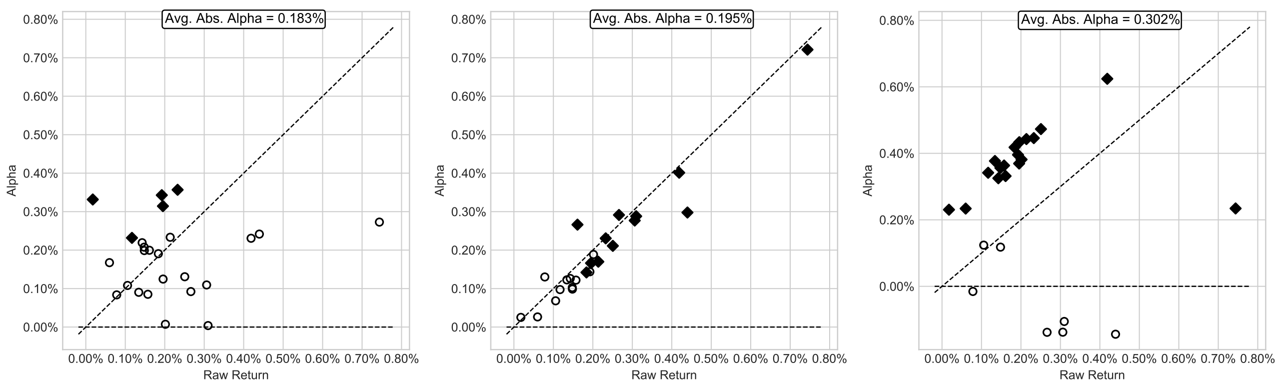


(a) `IPCA3` (b) `PCA7` (c) `IFF7`

Figure 3: **Alphas of double-sorted portfolios**

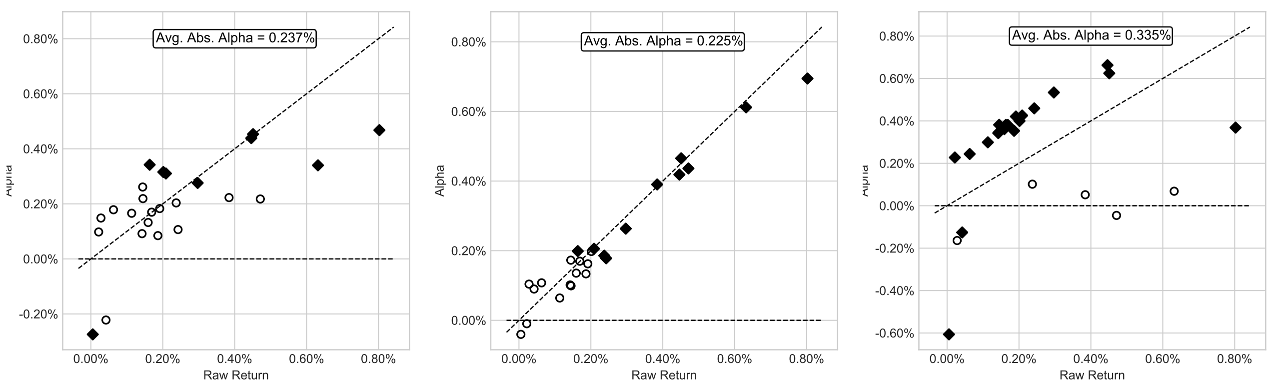
This figure shows the alphas of 25 portfolios sorted on `size` vs `r21_1`, `bm` or `max7` based on the set of factors from the restricted ($\Gamma_\alpha = 0$) IPCA model with $K = 4$ factors, a static principal component analysis with $K = 7$ (`PCA7`), and an instrumented factor model with $K = 7$ observable factor portfolios (`IFF7`). For the `IPCA4` and `IFF7` models the alphas are computed as the time-series average of the period-by-period portfolios residuals. For the static `PCA7` the alphas are computed as intercepts from time series regressions of portfolio returns on the latent or observable factors. All portfolios are re-leveraged to yield 5% daily volatility, consistent with the historical volatility of long-short portfolios. Significant alphas with absolute values of t-statistics greater than 2.0 are depicted with filled diamonds, while insignificant alphas are denoted with unfilled circles. All reported values are daily and expressed in percentage.

Panel A: Sorting on `size` and `r21_1`



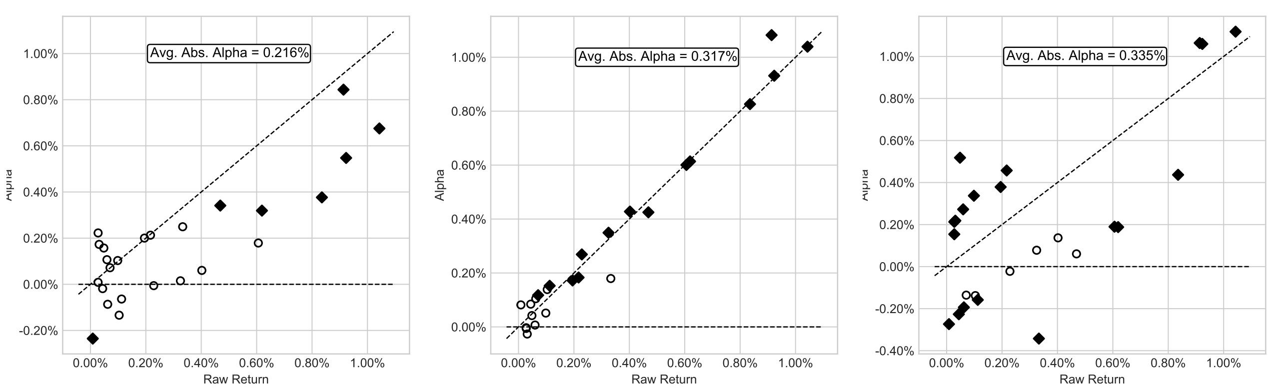
(a) `IPCA4` (b) `PCA7` (c) `IF7`

Panel B: Sorting on `size` and `bm`



(d) `IPCA4` (e) `PCA7` (f) `IF7`

Panel C: Sorting on `size` and `max7`

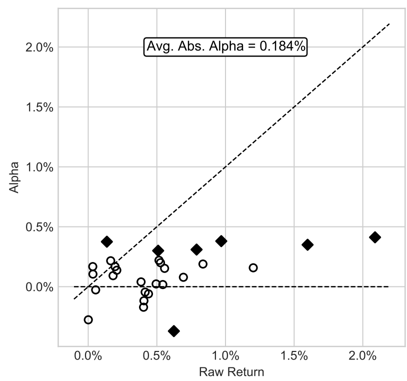


(g) `IPCA4` (h) `PCA7` (i) `IF7`

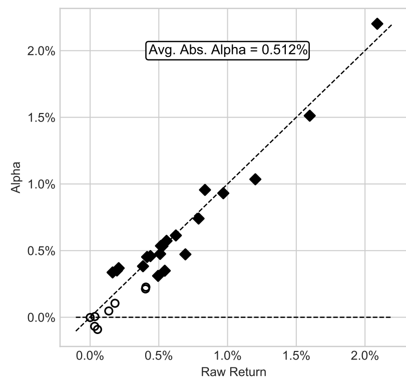
Figure 4: **Alphas of managed portfolios across sub-samples**

This figure shows the alphas of managed portfolios based on the set of factors from the restricted ($\Gamma_\alpha = 0$) IPCA model with $K = 4$ factors, a static principal component analysis with $K = 7$ (`PCA7`), and an instrumented factor model with $K = 7$ observable factor portfolios (`IFF7`). Panel A shows the results for the first sub-sample from September 1st 2017 to March 1st 2020. Panel B shows the results for the second sub-sample from March 2nd 2020 to September 1st 2022. For the `IPCA4` and `IFF7` models the alphas are computed as the time-series average of the period-by-period portfolios residuals. For the static `PCA7` the alphas are computed as intercepts from time series regressions of portfolio returns on the latent or observable factors. All portfolios are re-leveraged to yield 5% daily volatility, consistent with the historical volatility of long-short portfolios. Significant alphas with absolute values of t-statistics greater than 2.0 are depicted with filled diamonds, while insignificant alphas are denoted with unfilled circles. All reported values are daily and expressed in percentage.

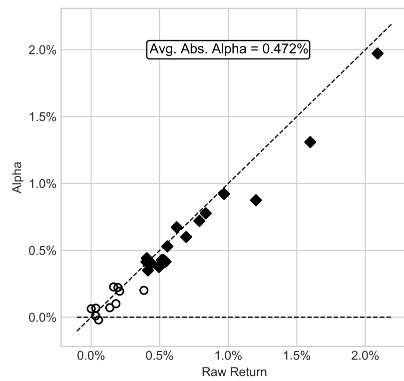
Panel A: First sub-sample



(a) `IPCA4`

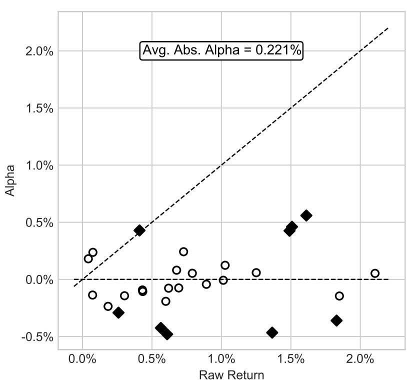


(b) `PCA7`

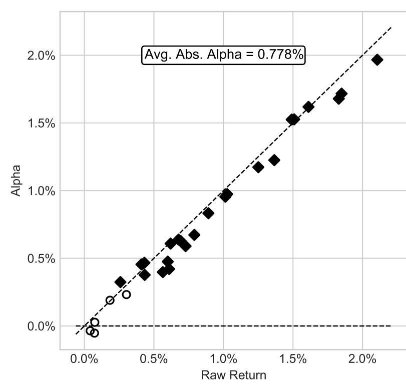


(c) `IFF7`

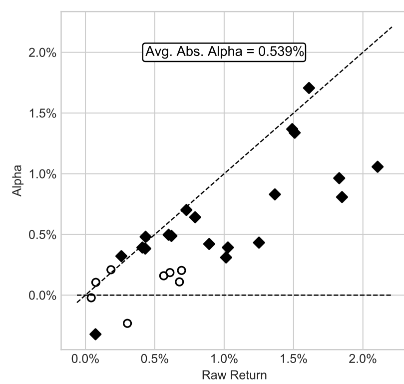
Panel B: Second sub-sample



(d) `IPCA4`



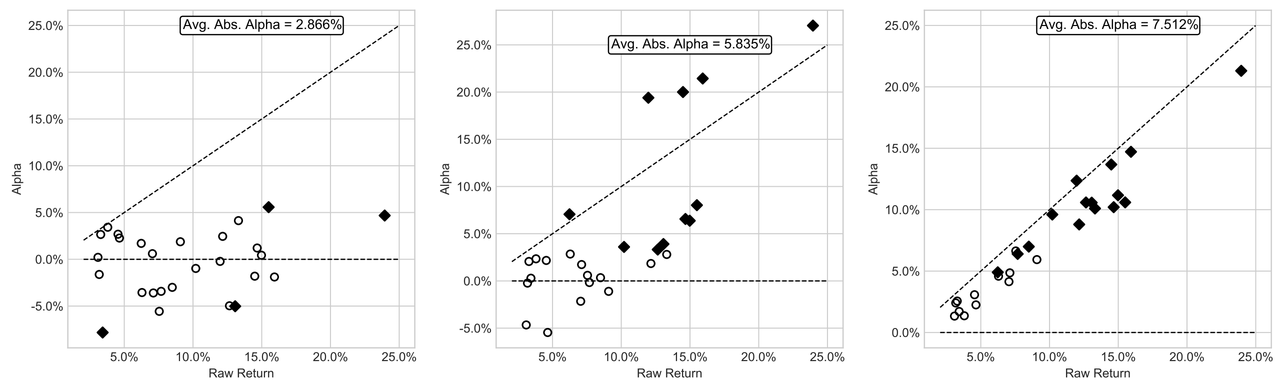
(e) `PCA7`



(f) `IFF7`

Figure 5: **Alphas of managed portfolios based on weekly returns**

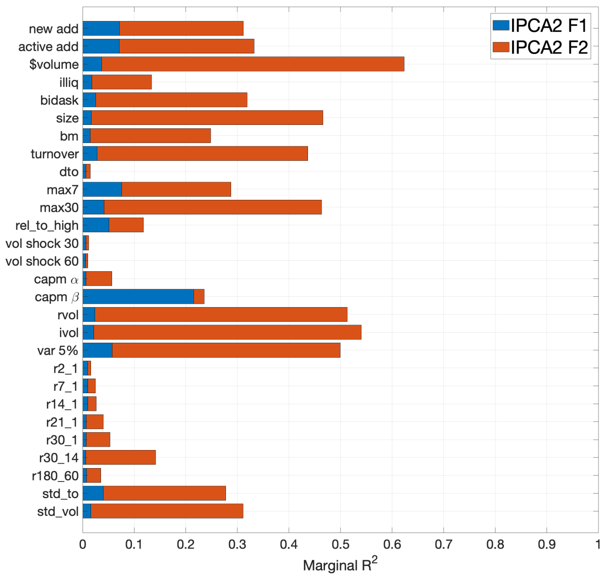
This figure shows the alphas of managed portfolios for the weekly returns. The alphas are calculated from the restricted ($\Gamma_\alpha = 0$) IPCA model with $K = 4$ factors, a static principal component analysis with $K = 7$ (`PCA7`), and an instrumented factor model with $K = 7$ observable factor portfolios (`IFF7`). For the `IPCA4` and `IFF7` models the alphas are computed as the time-series average of the period-by-period portfolios residuals. For the static `PCA7` the alphas are computed as intercepts from time series regressions of portfolio returns on the latent or observable factors. All portfolios are re-leveraged to yield 25% weekly volatility, consistent with the historical volatility of long-short portfolios. Significant alphas with absolute values of t-statistics greater than 2.0 are depicted with filled diamonds, while insignificant alphas are denoted with unfilled circles. All reported values are weekly and expressed in percentage.



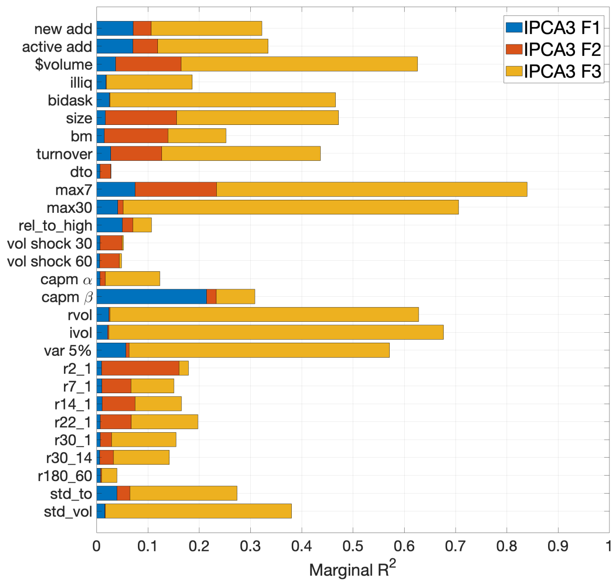
(a) `IPCA4` (b) `PCA7` (c) `IFF7`

Figure 6: **Marginal R^2 for IPCA factors**

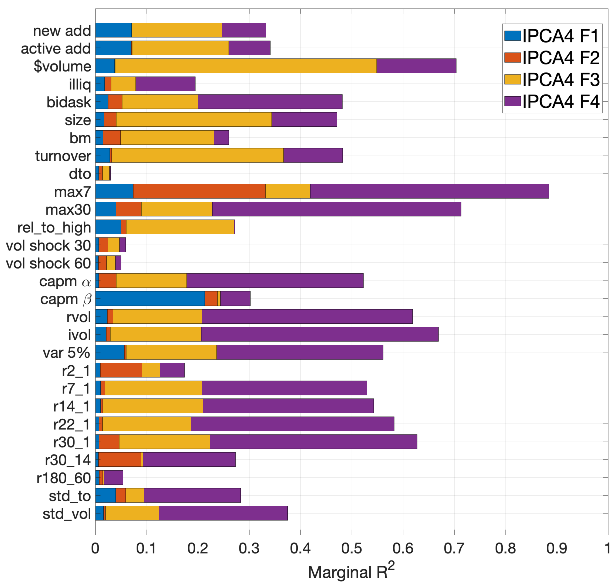
This figure shows the marginal R^2 of a set of auxiliary regressions in which the dependent variable is a given latent factor from an IPCA model, and the independent variables are the estimated managed portfolios for all 28 characteristics. The figure reports the results for different IPCA specifications with $K = 2, 3, 4, 5$ latent factors.



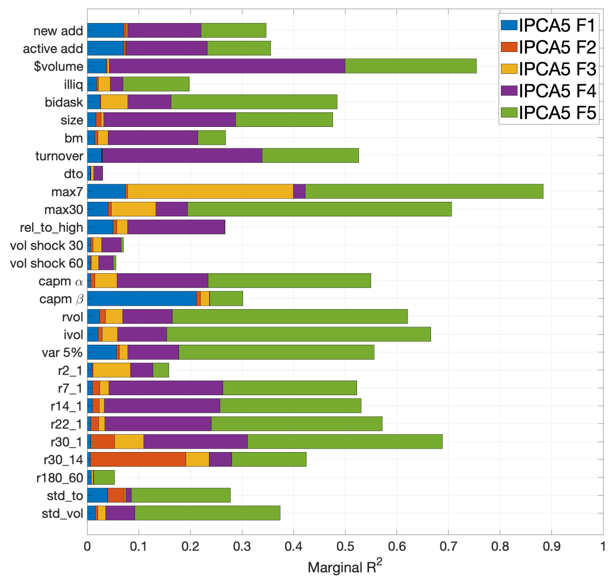
(a) `IPCA2`



(b) `IPCA3`



(c) `IPCA4`



(d) `IPCA5`

Internet Appendix to

A risk-based explanation of cryptocurrency returns

A Data cleaning

This Appendix describes the procedure related to sourcing, cleaning, and preparing the cryptocurrency database. The main results of the paper rely on two cryptocurrency databases.

A.1 Original sources

1. The CryptoCompare database is used to download aggregated and exchange level OHLC pricing and volume cryptocurrency data each day, where a day is defined with a start time of 00:00:00 UTC. We set tryConversion to 'true' and the tsym parameter to 'USD' for the regression and aggregated data-based portfolio sorts. We set tsym to 'USDT' and tryConversion to 'false' for the exchange-level portfolio sort robustness results.
2. The IntoTheBlock.com database, which is used to source information on blockchain activity, such as the number of new addresses and the number of active unique addresses. The day 'start time' is also set to be exactly 00:00:00 UTC.

A.2 Data pre-processing

We only retain cryptocurrency pairs if they have all available data from CryptoCompare and CoinGecko after merging. We consider a variety of pre-processing steps for a cryptocurrency to be included in the sample:

1. `Non-zero price and volume`: we exclude any pair that had zero traded volume or a zero price for any day t.
2. `Volume-to-market-capitalization`: we compute, for each pair and day t, the ratio of cryptocurrency traded volume to market capitalization and exclude any pair with a ratio > 1. This is a simple filter to screen out pairs with 'erroneous' or 'fake' volume. The measure is conservative - the median of the ratio is 0.001. This allows us to exclude any data points that are clearly errors.
3. `Cryptocurrency type`: We utilize cryptocurrency classification data from CoinMarketCap and screen out all cryptocurrencies which:
   - Are linked, are by backed or track the price of gold or any precious metal.
   - So-called 'wrapped' coins - i.e. WBTC.
   - Stablecoins, including those which are centralized (USDT, USDC) and algorithmically stabilized (DAI, UST) for all fiat currencies.
   - Centralized exchange based coins which are derivatives.
4. `API issues and suspicious trading activity`: As far as suspicious trading activity is concerned, a series of filter are implemented by CryptoCompare.com to mitigate the effect of suspicious trading activity: first, trade outliers are automatically excluded from the calculation of trading volume and therefore from the volume-weighting scheme. For a trade to be considered an outlier, it must deviate significantly either from the median of the set of exchanges, or from the previous aggregate price.[16]

[16] Such deviations can occur for a number of reasons, such as extremely low liquidity on a particular pair, erroneous

Second, exchanges are reviewed on a regular basis for each given cryptocurrency pair. Constituent exchanges are excluded if (1) posted prices are too volatile compared to market average of a given pair, (2) trading has been suspended by the exchange on a given day, (3) verified user or social media reports false data provision, or (4) malfunctioning of their public API. These steps mitigate the effect of fake volume and substantially reduces the exposure of the empirical analysis to concerns of misreporting of trading activity for some exchanges.[17]

A.3 Final sample

After all filters and checks we are left with an unbalanced panel of 332 cryptocurrency pairs which span the period from September 1st, 2017 to September 1st, 2022. As shown in Figure 1, the sample cover a fraction of the total market capitalization in the range of 70% to 95% of the total market value.

B Cryptocurrency characteristics

This section details the construction of variables we use in the main body of the paper and the relevant references. Unless otherwise specified we use the data sources outlined in Section A.

`new add.` : The number of unique addresses that appeared for the first time in a transaction of the native coin in the network. Liu et al. (2021) provide some preliminary evidence on the predictive content of new addresses for cryptocurrency returns.

`active add.` : The number of unique addresses that were active in the network either as a sender or receiver. Only addresses that were active in successful transactions are counted. As highlighted by Pagnotta and Buraschi (2018) such statistics approximate the network growth and the adoption base for a given cryptocurrency.

`bm.` : The "network-to-market value". This is calculated as the cumulative number of unique addresses over the current available supply times the current USD price. The current supply is the number of coins or tokens that have been mined or generated and corresponds to the number that are currently in public and company hands, which are circulating in the market and/or locked/vested. As suggested in Pagnotta and Buraschi (2018), the number of unique addresses represents a proxy for the fundamental value of a cryptocurrency. By dividing such value over the actual market value one can obtain a crude approximation of a valuation ratio.

`$volume.` : The total dollar amount of native tokens transferred across wallets within and across centralised exchanges.

`size.` : The market capitalization is defined as the product of the current available supply times the current USD price (see, e.g., Liu et al., 2022). The current supply is the number of coins or tokens that have been mined or generated and corresponds to the number that are currently in public and company hands, which

data from an exchange and the incorrect mapping of a pair in the API.

[17] Two additional comments are in order. First, notice that "fake" trading typically takes place on crypto-to-crypto trading on single, possibly small, exchanges which inflates trading volume in order to attract Initial Coin Offering's (ICO) listings and/or to manipulate the market (see, e.g., Li et al., 2018). By considering trading against a fiat currency and an aggregation over a large cross-section of exchanges, the risk that manipulation on a single exchange could affect the overall market activity is substantially mitigated. Second, the fact that we focus on transactions that take place on regular trading exchanges should mitigate the concern that market activity is primarily driven by illegal activities. The latter typically do not take place on registered centralised or decentralised exchanges but through peer-to-peer transactions on the blockchain (see Foley et al., 2019 and Griffin and Shams, 2020).

are circulating in the market and/or locked/vested. This definition follows the blueprint in Fama and French (1993).

`rvol.` : We follow Yang and Zhang (2000) and calculate the daily realized volatility calculated based on daily OHLC prices.

`bid-ask.` : The bid-ask spread is the average of two alternative synthetic approximations based on OHLC prices by Abdi and Ranaldo (2017) and Corwin and Schultz (2012). On a given day and for a given cryptocurrency pair we calculate both proxies and take the simple average between the two.

`illiq.` : We follow Amihud (2002) and calculate a price impact (illiquidity) measure as the ratio between the absolute value of the cumulative intraday returns and the aggregate daily trading volume expressed in $.

`capm` β. : The market beta is calculated based on a 30-day rolling window. We follow Lewellen and Nagel (2006) and calculate the beta as the sum of the coefficients of daily returns on the market excess return and one lag of the market excess returns. The market portfolio is calculated as the value-weighted average of the asset returns available on each day t.

`capm` α. : The intercept from a CAPM regression calculated based on a 30-day rolling window (see description of the `capm` β).

`ivol.` : The standard deviation from the residuals from a CAPM regression calculated based on a 30-day rolling window (see description of the `capm` β).

`turnover.` : Turnover is last day's trading volume in $ over the current supply (see Datar et al., 1998). The current supply is the number of coins or tokens that have been mined or generated and corresponds to the number that are currently in public and company hands, which are circulating in the market and/or locked/vested.

`dto.` : We follow the logic in Garfinkel (2009) and define de-trended turnover as the ratio of daily volume in $ to current available supply, minus the daily market turnover and de-trend it by its 180 trading day median. The daily market turnover is a value-weighed aggregation of the individual assets' turnover.

`std_to.` : The standard deviation of the residuals from a 30-day rolling window regression of daily turnover on a constant (see Chordia et al., 2001).

`std_vol.` : The standard deviation of the residuals from a regression of daily trading volume on a constant (see Chordia et al., 2001).

`rel_to_high.` : Closeness to 90-day high is the ratio of the cryptocurrency price at the end of the previous day and the previous 90-day high. This adapts to a shorter time span the logic in George and Hwang (2004).

`max.` : Maximum daily return in the previous month following Bali et al. (2011).

`vol shock` l`d.` : We follow Llorente et al. (2002) and construct the log deviation of trading volume from its trend estimated over a rolling period of $l = 30, 60$ days. The log standard deviation computed over the same rolling window is used to standardise the estimates due to cross-sectional imbalances (see Babiak et al., 2022).

`r2_1.` : Short-term reversal as in Jegadeesh (1990)

`r`l`_2.` : We follow Liu et al. (2022) and construct a variety of momentum strategies based on the cumulative return from $l = 7, 13, 22,$ and 31 days before the return prediction to two days before.

`r30_14.` : We define intermediate momentum as the cumulative returns from 30 days before prediction to 14 days before. This is an adaption on a higher frequency time span from Novy-Marx (2012).

`r180_60.` : We define long-term reversal is the cumulative return from 180 days before the return prediction to 60 days before. This is an adaption of De Bondt and Thaler (1985) to a higher frequency setting.

`VaR(5%).` : The historical Value-at-Risk at 5% calculated based on past 90-day returns.

C Additional results

This section provides a series of additional results. We first look at the performance of an IPCA with and without restriction on the intercept parameters. Second, we look at the ability of IPCA latent factors extracted from the cross section of cryptocurrency returns to price 25 equity portfolios sorted on size and book-to-market, or size and momentum. Third, we look at the correlation between the latent factors extracted from an IPCA and a static PCA model.

C.1 IPCA performance with and without intercept

Table C1 reports the in-sample asset pricing performance of different IPCA specifications with and without restrictions on the intercept coefficients Γ_α. In addition, we report the values of Bai and Ng (2002) information criteria for each specification of latent factor models as well as the p-values for the test of $\Gamma_\alpha = 0$ for IPCA based on a wild bootstrap with 10,000 draws.

The results suggest that the explanatory power of the characteristic-driven intercept is limited, that is the IPCA explains the variation in realised and expected returns solely based on risk compensation (see Kelly et al., 2019). For instance, the R^2_{tot} from a baseline `IPCA4` assuming $\Gamma_\alpha \neq 0$ is equal to 13.5% versus 13.4% when the matrix of coefficients Γ_α is restricted to zero. That is, the additional variation that is captured by the intercept is minimal with respect to the latent factors. The spread in the performance of the single-factor specification remains negligible for the R^2_{pred} metric. Increasing the number of latent factors does not lead to a widening of the gap between the restricted and unrestricted IPCA specifications.

The same results hold when we use the set of 28 managed portfolios as test assets. For instance, the $R^2_{tot,x}$ ($R^2_{pred,x}$) is 50.5% (2.05%) for the unrestricted IPCA versus a 51.9% (1.93%) for the restricted IPCA. Interestingly, the relative pricing error is largely in favour of the restricted IPCA specification. For instance, the RPE from the `IPCA4` restricted is 6.7% against a 96% RPE from the equivalent unrestricted IPCA model.

C.2 Pricing equity with IPCA cryptocurrency factors

Figure C1 reports the results from a series of time-series regressions in which each test asset is regressed either on the Fama-French `FF5` model (left panels) or the three latent factors from the crypto `IPCA4` (right panels). The top panels report the average absolute pricing error by using 25 portfolios sorted on size and book-to-market as test assets. The average absolute pricing error from the `FF5` model is 3.4% annualized against a 29.5% obtained from cryptocurrency factors. In addition, for the `IPCA3` model the alphas are clustered around

Table C1: **Asset pricing performance with and without intercept**

This table reports the in-sample asset pricing performance of different IPCA specifications with and without restrictions on the intercept coefficients Γ_α. In addition, we report the values of Bai and Ng (2002) information criteria for each specification of latent factor models as well as the p-values for the test of $\Gamma_\alpha = 0$ for IPCA based on a wild bootstrap with 10,000 draws. The data are sampled daily from September 1st 2017 to September 1st 2022, where a day is defined with a start time of 00:00:00 UTC. Daily prices and volume are aggregated across more than 80 different centralised exchanges.

	$\Gamma_\alpha \neq 0$							$\Gamma_\alpha = 0$						
	IPCA1	IPCA2	IPCA3	IPCA4	IPCA5	IPCA6	IPCA7	IPCA1	IPCA2	IPCA3	IPCA4	IPCA5	IPCA6	IPCA7
R^2_{tot} (%)	11.42	12.29	12.91	13.54	14.06	14.58	15.15	11.07	12.01	12.74	13.37	13.98	14.50	15.08
R^2_{pred} (%)	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.00	0.15	0.31	0.32	0.33	0.33	0.33
R^2_{ts} (%)	18.65	18.94	19.11	19.31	19.41	19.62	19.69	18.61	18.89	19.14	19.39	19.51	19.63	19.69
R^2_{cs} (%)	9.26	10.12	10.80	11.46	12.01	12.57	13.17	8.98	9.85	10.63	11.32	11.96	12.52	13.13
$R^2_{tot,x}$ (%)	18.87	38.77	42.56	50.48	53.99	56.69	58.53	16.76	33.72	43.36	51.91	53.57	56.51	58.38
$R^2_{pred,x}$ (%)	2.11	2.07	2.08	2.05	2.05	2.05	2.04	0.03	0.49	2.01	1.93	1.96	1.95	1.95
$R^2_{ts,x}$ (%)	10.47	31.78	35.74	43.67	47.26	50.14	52.09	8.18	26.42	36.53	45.16	46.91	49.90	51.94
$R^2_{cs,x}$ (%)	13.56	30.13	33.40	41.70	45.38	48.35	50.53	12.34	25.14	34.92	43.25	45.17	48.62	50.45
RPE	100.06	88.54	101.24	95.96	228.89	158.14	171.20	99.48	84.03	6.43	6.66	5.86	4.70	4.91
IC	-7.50	-7.66	-7.61	-7.64	-7.59	-7.54	-7.46	-7.47	-7.58	-7.62	-7.67	-7.58	-7.53	-7.46
$H_0: \Gamma_\alpha = 0$ (pval)	0.00	0.00	0.00	0.02	0.07	0.98	0.96							

the 45-degree line, which suggests the factors extracted from the panel of cryptocurrency returns do not provide significant pricing information for equity markets. Results are similar by looking at 25 portfolios sorted on size and momentum (bottom panels). The `FF5` equity factor model produces an average absolute pricing error equal to 4.9% annualized, against a sixfold higher pricing error of 30.7% obtained when using the `IPCA3` factors. Yet, none of the alphas from the `FF5` pricing model is significant.

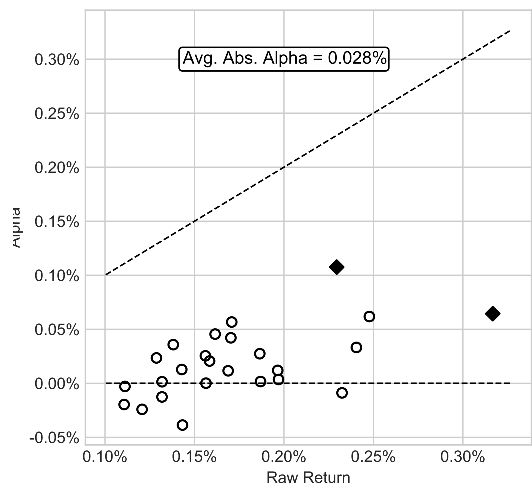
C.3 IPCA vs PCA factors

Since latent factors can be only identified up to a rotation, we assess the correlations between PCA and IPCA latent factors by a series of spanning regressions, that is we regress each of the factors from the baseline `IPCA3` model on all of the seven factors from the competing `PCA7`. The choice of these two models is consistent with the main results represented in the paper. Table C2 shows two interesting results. First, none of the static principal components perfectly correlates with the IPCA ones: the constant, meaning the unexplained factor returns, is strongly significant for all the three latent factors. Second, while the multiple correlation coefficients ($\sqrt{R^2}$) for the second IPCA factor shows a strong correlation of 84% with the PCA factors jointly, the first and third IPCA factors have a smaller multiple correlation of 43% and 63%, respectively. These results suggest that the factors extracted from a standard principal component analysis factors do not span the IPCA factors.

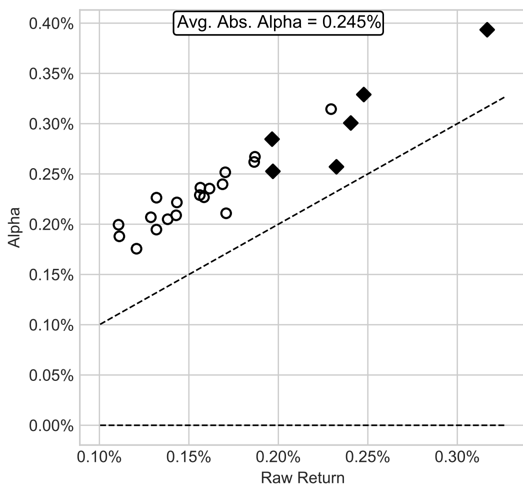
Figure C1: **Alphas of double-sorted equity portfolios**

This figure shows the alphas from a time series regression of double-sorted equity portfolios on the set of factors from either the restricted ($\Gamma_\alpha = 0$) IPCA model with $K = 4$ factors or the five-factor model of Fama and French (2015). The left (right) plots illustrate the results for the Fama-French (IPCA) model. The test assets are 25 equity portfolios sorted on (1) size and book-to-market (Panel A) and (2) size and momentum (Panel B). Significant alphas with absolute values of t-statistics greater than 2.0 are depicted with filled diamonds, while insignificant alphas are denoted with unfilled circles. All reported values are daily and expressed in percentage.

Panel A: Size and book-to-market

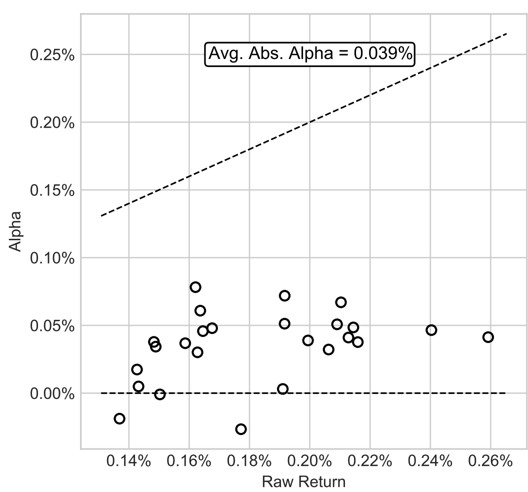


(a) `Fama-French five factors`

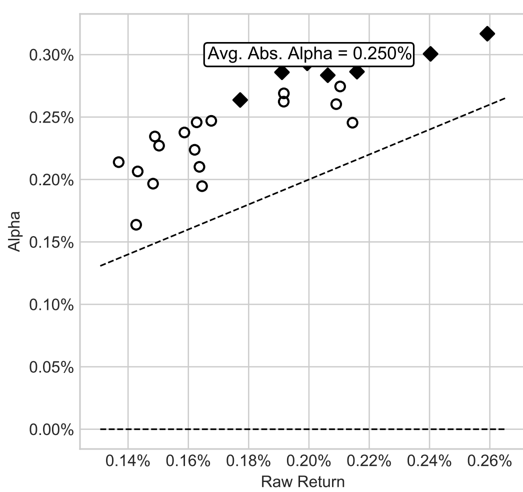


(b) `IPCA4`

Panel B: Size and momentum



(c) `Fama-French five factors`



(d) `IPCA4`

Table C2: **IPCA vs PCA spanning regressions**

This table reports the results of a number of time-series regressions in which we regress each latent factor from a restricted ($\Gamma_\alpha = 0$) IPCA model with $K = 3$ on six latent factors extracted from a static PCA method. We report the estimates and label with ***, **, * those coefficients significant at a 1%, 5%, and 10% confidence level, based on robust standard errors. The multiple correlation coefficients in the last row are measured as the square root of R^2. The sample factor returns is daily from September 1st 2017 to September 1st 2022.

	IPCA1		IPCA2		IPCA3	
$\alpha(\%)$	2.741	***	0.284	***	0.422	***
PCA1	-0.012	***	0.076	***	-0.041	***
PCA2	0.010		-0.003		-0.009	
PCA3	0.042	***	-0.023	*	-0.023	***
PCA4	0.022	*	0.031	***	-0.034	***
PCA5	0.003		0.013	***	0.005	
PCA6	0.043	***	-0.016	***	-0.026	***
PCA7	0.023	*	-0.036		-0.046	*
$\sqrt{R^2}$	0.431		0.841		0.633	