Domestication of Cryptoassets: Theory and Evidence

Dong-Hyun Ahn[a], Jaewon Choi[b], Kyu Ho Kang[c] and Seongdeok Ko[d] *

December 4, 2023

Abstract

This paper begins with a state-space analysis on a structural break in relationships between cryptoassets and traditional assets in terms of their returns. A Bayesian MCMC method detects a statistically significant break at the end of February, 2020, which roughly matches the time period when the institutionalization of cryptoassets became sufficiently noticeable. Based on this finding, we investigate an issue of developing a legitimate measure that can distinguish between domesticated cryptos and non-domesticated cryptos where 'domestication' refers to two characteristics of individual cryptos, one for comovment in price with stocks and the other for pricing consistency with the stocks. A natural candidate for this domestication measure would be an orthogonal extension in the stochastic discount factor when a cryptoasset is augmented as an additional basis asset to the stock basis assets. However, we find that this variation of the Hansen-Jagannathan distance measure results in an counter-intuitive inference about domestication and we adjust it by introducing what we dub 'amplification factor,' which is either linearly or quadratically proportional to idiosyncratic risks of the crypto. We devise a couple of novel empirical tests for horse racing among the candidate domestication measures and all of them signify that the seemingly mispriced value against the minimum-norm stochastic discount factor of the stock basis assets as a benchmark performs the best in terms of precision in sorting which of cryptos are domesticated. Based on this admissible domestication measure, we find that the proportion of domesticated cryptos has skyrocketed after the break from 7% to 36%. In addition, the stochastic discount factor governing the prices of *domesticated* cryptos is explicable by the representative stock market factors such as the Fama-French 5 factors and the q^5 factors, but not by the the crypto-specific factors proposed by Liu, Tsyvinski and Wu (2022). In contrast, the crypto-specific factors show predominant power in explaining the stochastic discount factor ruling the prices of *non-domesticated* cryptos whereas the stock-specific factors fail to do so. The combined results re-affirm that the domestication measure selected in the horse racing competition is truely reliable.

JEL classification: E32; E44; G12; G14
Keywords: Cryptoasset; Institutionalization, Domestication, Stochastic Discount Factor, Nonparametric, Bootstrapping, MCMC, State-space Analysis

---

*[a]Corresponding author. Department of Economics, Seoul National University, Building 16, Gwanak-ro, Gwanak-gu, Seoul 08826, Korea. e-mail: ahnd@snu.ac.kr; [b] University of Illinois Urbana-Champaign, 4035 BIF, Champaign, IL 61820. e-mail : jaewchoi@illinois.edu; [c]Department of Economics, Korea University, Political Science and Economics Building 328, Anam-ro, Sungbuk-Gu, Seoul 02841, Korea. e-mail: kyuho@korea.ac.kr; [d]Department of Economics, Seoul National University, Building 16, Gwanak-ro, Gwanak-gu, Seoul 08826, Korea. e-mail: ksdskd@snu.ac.kr. We wish to thank seminar participants at Samsung Global Research, Hanyang Unniversity and Korea University for their insightful suggestions and comments. The usual disclaimer applies.

1 Introduction

> "These walls are funny. First you hate 'em, then you get used to 'em. Enough time passes, you get so you depend on them. That's institutionalized."
>
> — The Shawshank Redemption

For the last decade or so, cryptoassets have emerged as one of the most controversial but phenomenal subjects among not only the general public but also investment professionals. It all started with the publication of a white paper entitled 'Bitcoin: A Peer-to-Peer Electronic Cash System' written by a yet-to-be-identified author under the pseudonym of Satoshi Nakamoto in October 2008, which was immediately followed by the release of the Bitcoin protocol as an open source software in January 2009. The explosive attention with Bitcoins, driven at first by cypherpunks and later by early individual adopters has been expanded to a much more mainstream form of investment. Such a furious pace of progress demonstrated by Bitcoin has inspired and sparked a plethora of supplies of other cryptoassets to such an extent that the aggregate market value of established cryptoassets on offer reached nearly \$3 trillion at the end of 2021, dwarfing the world's largest company, Apple, while amassing more than 200 million users all over the world.[1]

Despite their proliferation, the future prospects of the cryptoassets still remain as contentious, which splits the global investment community down the middle. Some ardent supporters argue that the game has only just begun: given the durability and scalability of the cryptoasset networks, their potential outside the areas of digital assets and finance is yet to be fully explored. In contrast, skeptics have raised doubts on their underlying technological and economic concept, even criticizing them as a gimmick with all the markings of a Ponzi scheme.[2]

Love them or hate them, given the size, it has grown simply too big to ignore, and a stampede of institutional professions for these digital assets ensued.[3] Traditional institutional players, with Fidelity, Goldman Sachs, JP Morgan, Black Rocks, CME, and CBOE are but a few examples. Institutional investors prefer investment via financial intermediaries rather than opting to hold cryptoassets directly for several considerations and investment funds have emerged as a natural gateway for investors to obtain cryptoasset exposure.[4] Auer et al. (2022), based on a wide range of data sources, document that flows into closed-end funds centering upon cryptoassets slowly and steadily have grown before a sudden and dramatic surge observed during 2019. Specifically, cumulative net inflows into investment funds has skyrocketed from about \$1 billion at the end of 2018 to \$7 billion in the beginning of 2020. As a result, the total assets under management are as large as \$30 billion at the end of 2021 after once reaching \$60 billion over the course of that year. It is true that this amount still represents a small fraction of the entire cryptoasset market, but the presence of institutional professions is remarkable given the fact that this market is still in a nascent stage. Simply put, the 'institutionalization' of the cryptoassets is underway on a large scale.

Active participation of institutional investors, once it reaches a certain scale, may cause a change in the price behavior of a cryptoasset. First, the institutionalization of the cryptoasset may induce its

[1] See Auer et al. (2022).

[2] See, for example, Taleb (2020), Roubini (2021), and Acemoglu (2021) among many others.

[3] See Fidelity (2021), Street (2021) and OECD (2022), among many others.

[4] Auer et al. (2022) document that these considerations include operational complexity, custodian safety and a difference in accounting rules regarding unrealized capital gains on crytpoassets and investment funds.

price movement in tandem with the traditional assets classes. Second, if so, its price may become subject to the 'pricing rule' prevailing in those asset classes. In this paper, if a particular cryptoasset satisfies the above two conditions, we claim that it is 'domesticated.' Thus the definition of domestication is built upon two building blocks, the cryptoasset's comovement in price with a mainstream asset class and its price alignment with that asset class. Based on this definition, we investigate which cryptoasset has been domesticated and also whether the percentage of domesticated cryptoassets has significantly increased in conjunction with the aforementioned institutionalization process.

With respect to comovement in price, unlike the traditional financial assets, most of cryptoassets do not accompany any future cash flows.[5] Therefore, there is no reason for the cryptoasset prices to move 'fundamentally' in tandem with those of the traditional assets. However, the absence of comovement in fundamentals does not rule out the possibility of price comovement. Barberis, Shleifer and Wurgler (2005) argue that in economies with frictions and/or with irrational investors, comovement in prices is delinked from comovement in fundamentals. They call this non-fundamentally driven comovement as the "friction-based" and/or "sentiment-based" theories of comovement.

Barberis, Shleifer and Wurgler (2005) further subdivide them into three specific views: the category view, the habitat view and the information diffusion view. The category view, which is proposed by Barberis and Shleifer (2003) states that some investors allocate funds at the level of certain categories rather than at the individual asset level. If some of the investors using categories share correlated sentiment, and if their trading affects prices, then as they move funds from one category to another, their coordinated demand induces common factors in the returns of assets in the same category. The habitat view is based on the observation that many investors trade on their own preferred habitats, i.e., only a subset of all available securities. When these investors alter their exposure to the securities in their habitat, a common factor in the returns of these securities is induced. Finally, the information diffusion view holds that information is incorporated more quickly into the prices of some assets than others, primarily due to some market frictions. Among the three views, because the cryptoassets do not generate any future cash flows, they do not share any information with the traditional assets and it is hard to believe that the information diffusion is behind comovement in price. The category view and the habitat view are then the only legitimate candidates for price comovement of the cryptoassets with other asset classes, if anything. However, these two views, despite their subtle difference in theoretical arguments, are difficult to distinguish in an empirical analysis. For that reason, the terms 'category' and 'habitat' are used interchangeably in Barberis, Shleifer and Wurgler (2005). Following them, we combine the two views as the 'category view' in our analysis. As such, our analysis can shed a new light on the driving force behind 'comovement' proposed by Barberis, Shleifer and Wurgler (2005) by testing it in a novel and clean setup.

To begin with, we collect the individual cryptoasset data from CoinMarketCap API over the period from January 2017 to March 2022, following Liu, Tsyvinski and Wu (2022).[6] Using the biweekly frequency data, we investigate comovement between Bitcoin and other relevant traditional assets by employing two alternative tests. In the first test, we consider gold, dollar and S&P500 as traditional assets against Bitcoin (and MVIS CryptoCompare Digital Asset 100 Index (MVDA), a market cap-weighted index that tracks the performance of the 100 largest cryptoassets). We

[5] There exist some ways by which one can earn passive income similar to dividends in the crypto space by staking or 'HODLing' a cryptocurrency. However, strictly speaking, they are more or less equivalent to stock lending income.

[6] We also re-examine the entire analysis with the alternative dataset collected from the Cryptocompare API. The estimation results are not only qualitatively but also quantitatively very similar.

examine a state-space econometric model in which there is a common factor along with asset-specific factors (e.g., gold-specific, dollar-specific, S&P500-specific and Bitcoin-specific factors). Therein, Bitcoin returns are designed to be potentially affected by all the factors in addition to its own Bitcoin-specific factor. More importantly, this state-space model is equipped with a two-state Markov chain such that a structural break is endogenously determined. Using the Bayesian MCMC (Markov chain Monte Carlo) method, we estimate all the model parameters and state variables along with the structural break point. The estimation results are striking! They designate a statistically very significant break on February 21, 2020. In addition, the variance decomposition indicates that before the break, the returns on Bitcoin were alienated from other assets and almost fully (95.6%) explained only by Bitcoin itself, which delivers conclusive evidence that the Bitcoin market was segmented from the traditional assets. In contrast, after the break, the majority of its returns, 62.7%, is explained by the S&P500 returns whereas the self-explanatory portion of its variance has remarkably shrunken to as small as 24.9%. Even after the structural break, the Bitcoin returns are not significantly related with dollar and gold returns. Therefore, the claim that Bitcoin is an alternative digital currency to traditional fiat money systems is statistically rejected. And similarly, the claim of Bitcoin as 'digital gold' is also equally untenable.

One thing to notice among these results is the breakpoint, February 2020. This is precisely the time period when the cumulative net inflows into investment funds, which began to gain the momentum in the beginning of 2019, reached the top according to the study of Auer et al. (2022).[7] Consequently, the breakpoint we identified can be thought of as the date when the institutionalization of Bitcoin gained muscle mass enough to make a statistically significant change in the price comovement of Bitcoin with stocks, the representative risky assets. In addition, the empirical finding that Bitcoin 'comoves' in price with stocks is in line with the survey of PwC (2022). It documents that about 38% of traditional hedge funds invest in cryptoassets in the first half of 2022. More importantly, among the most popular hedge fund strategies that invest in digital assets include multi-strategy (32%), macro (21%), equity (18%) and systematic (12%). These funds are all exposed to equities in their investment. Therefore, we can surmise that after the break, some cryptoassets such as Bitcoin became to be inserted into the same 'category' with stocks, which induces non-fundamentally driven comovement as suggested by Barberis, Shleifer and Wurgler (2005).

The second study on comovement is to utilize the fact that the cryptoassets trade around the clock across the globe. Using this unique feature of the cryptoassets in their trading hours, we explore whether the overnight returns on Bitcoin affect the overnight returns on the S&P500 index (or Nasdaq) from the closing price of the previous trading day to the opening price on the following business day. In testing this relationship, we include the previous day's intraday returns on S&P500 to control for a potential of autocorrelation in S&P500.[8] In addition, given the fact that the trading hours of the U.K.'s stock market partially overlaps with the overnight closing of S&P500, we further include the FTSE100 index return to control for the spill-over effect of inter-market stock prices. We estimate this structural regression in the presence of stochastic volatility where the spill-over is also allowed in the second moment, i.e., volatility. The estimation result strongly backs up the empirical findings of the first test. Before the break, we find no evidence of spill-over effect of Bitcoin on the stocks. In contrast, after the break, the overnight return of Bitcoin affects the overnight returns on the S&P500 with statistical significance at 1%. This statistical significance is maintained even when we control for the FTSE 100 returns. We also conduct a similar analysis

[7] See the left panel of Graph 2 therein.

[8] Some investment professions believe that trading patterns and trends on the intraday will influence what the market will experience after its close.

with the weekend returns, which are a special part of the overnight returns in order to see whether the spill-over effect is still alive over the longer time span during which the U.S. stock market sleeps. The test ends up with a similar result; the spill-over effect of Bitcoin on the S&P500 during the weekend is statistically significant. Combined with the state-space analysis, these test results signify the post-break comovement in price between Bitcoin and S&P500 Index.

Given the empirical finding that the price of Bitcoin began to move in tandem with stocks, a natural question then arises as to whether the cryptoasset price is also aligned with the pricing rule that stocks are subject to. This is the second building block of 'domestication' defined above. For that purpose, one may consider the integration measures proposed in extant literature. Since pricing consistency between the two markets is central in measuring integration, a benchmark pricing model is a necessary and central input, and in thar regard, there are two alternative approaches. The first approach, which has been popularly adopted in the international asset pricing literature, is to evaluate integration based on on a parametric, but potentially misspecified, asset pricing model with some pre-specified factors.[9] The second approach is nonparametric and relies upon a stochastic discount factor (SDF from hereon) as a pricing benchmark. The representative measures developed in this line of approach include Chen and Knez (1995) and Bekaert and Urias (1996) among others.

Since the cryptoassets differ from stocks in fundaments, some extant asset pricing models that show strong performance in explaining cross-sectional dispersion of stock returns may not be applicable. For example, the five factor model of Fama and French (2015) is composed of the factor-mimicking portfolios, which are built upon accounting variables such as book-to-market ratio, operating profitability and investment ratio. Similarly, the original q-factor model Hou, Xu and Zhang (2015) and its augmented version, the q^5 factor model of Hou et al. (2021) also need information regarding investment ratio, profitability and expected growth. All of those accounting-based firm characteristics are not available for cryptoassets. In addition, even in the case that there are some available asset pricing models, the resulting inference in the space of cryptoassets might be unacceptably sensitive to the choice of the benchmark pricing model. With this concern, we opt for the second approach as a natural choice in this paper.

However, domestication is different from integration, albeit closely related. Domestication refers to a relation of a certain individual asset to a market whereas integration refers to a relation of a market vis-a-vis another market. The reason this paper addresses domestication as opposed to integration is twofold. Firstly, the number of cryptoassets varies even on a daily basis because screening process of cryptoassets for listing on coin exchanges is much simpler due to the absence of listing standards as strict as those imposed on stock exchanges.[10] The wild fluctuations in the number of cryptoassets make it difficult to construct a set of well-diversified portfolios of cryptoassets that can well approximate and represent the attainable set of the cryptoasset market.[11]

Secondly and related, it is hard to believe that the cryptoasset market as a whole is mature enough to be ready for a test on whether it is integrated with the stock market. As of yet, the institution-

[9] See, among many others, Stulz (1981), Adler and Dumas (1983), Cho, Eun and Senbet (1986), Korajczyk and Viallet (1989) and Korajczyk (1996).

[10] The major U.S. stock exchanges are equipped with strict listing standards that encompass the rules on corporate governance and audit committees, under the supervision of the Securities and Exchange Commission.

[11] Fama and French (2015) argue that a set of poorly diversified portfolios have low power in tests of asset pricing models. In fact, such a set also results in a large 'size' problem in the tests as well. In addition, if the set does not sufficiently approximate the attainable set of the cryptoasset market, it under-represents it and yields a large amount of inference errors in diagnosing the integration of two markets.

alization is still in its early stage. As such, we focus on whether a certain individual cryptoasset is domesticated enough to meet the minimum requirements for admissibility in the pricing rules governing the stocks.

A natural candidate measure would be the distance measures suggested in a series of seminal studies by Hansen and Jagannathan. To begin with, we construct a global minimum-norm SDF (m_x^*) from a set of stock basis assets, which are the Fama-French 25 portfolios sorted on book-to-market and size coupled with the locally risk-free asset in our analysis. Then we add a certain cryptoasset to the set of basis assets and, from this augmented set of basis assets, we re-construct the two minimum-norm SDFs, one with no restriction (referred to as $m_{x_a}^*$) and the other with the restriction that its mean is the same to that of m_x^* (referred to as $m_{x_a}^*|\mu_m$). Then, we measure a distance between $m_{x_a}^*$ (or $m_{x_a}^*|\mu_m$) and m_x^*, where the distance is measured by the second norm of the difference between the two SDFs, i.e., $\|m_{x_a}^* - m_x^*\|$ (or $\|m_{x_a}^*|\mu_m - m_x^*\|$). $\|m_{x_a}^*|\mu_m - m_x^*\|$ is equivalent to the amount of the upward shift in the Hansen-Jagannathan bound at the mean of m_x^* ($= \mu_m$) whereas $\|m_{x_a}^* - m_x^*\|$ is tantamoun to the Hansen-Jagannathan distance that is modified such that a thoeretical SDF in the original Hansen-Jagannathan distance is replaced by the nonparametric SDF, m_x^*.[12]

Through an extensive analysis, we show that these two distance measures can be re-written as the ratio of the absolute pricing error of the cryptoasset to its idiosyncratic risk, where the pricing error (α_c^*) is based on using m_x^* as the relevant benchmark SDF. The idiosyncratic risk is the replication error, which is the second norm of residuals from a regression of the cryptoasset's payoff against the original set of basis assets. Specifically, the second-norm of residuals from the regression 'without' intercept ($\|\omega_c\|$) is the idiosyncratic risk relevant for $\|m_{x_a}^* - m_x^*\|$ while that of residuals from the regression 'with' intercept ($\|e\|$) is the idiosyncratic risk corresponding to $\|m_{x_a}^*|\mu_m - \mu_x^*\|$. As such, the two distance measures, $\frac{|\alpha_c^*|}{\|\omega_c\|}$ and $\frac{|\alpha_c^*|}{\|e_c\|}$ can be interpreted as 'normalized' absolute pricing errors where normalization is made by the idiosyncratic risks. Accordingly, the cryptoassets which move less in tandem with the stocks are more likely to be sorted as domesticated, which is at odds with the first property required for domestication, i.e., comovment! To make matters worse, the pricing errors are empirically shown to be strongly correlation with the idiosyncratic risks so that the large pricing errors are inclined to be neutralized by their corresponding idiosyncratic risks. To remedy this counter-intuitive problem innate in the distance measures, we introduce an amplification factor, which is either linearly ($\|\omega_c\|$ or $\|e\|$) or quadratically ($\|\omega_c\|^2$ or $\|e\|^2$) proportional to the idiosyncratic risks and adjust the distance measures by multiplying them by the amplification factor. As a consequence, for each distance measure, we can consider three alternative candidate domestication measures depending on the magnitude of the amplification factor, but one candidate measure is found to be shared by the two distance measures so that, all told, we end up with five candidate measures. Then a remaining question is which candidate measures are more valid, i,e., more correct in distinguishing between domesticated cryptoassets and non-domesticated ones.

To answer this question, we come up with a two-step estimation procedure. In the first step, we apply each candidate measure to every individual stock available in the CRSP (the Center for Research in Security Prices) and construct a cross-sectional distribution. Because this distribution is composed of all available individual stocks, which are fully domesticated by construction, it is a natural benchmark distribution under the null that an individual asset is domesticated. In the

[12] The Hansen-Jagannathan bound (HJ bound from here on), which is derived in Hansen and Jagannathan (1991) is a lower bound on the second norm of the SDF as a function of its mean. In contrast, the Hansen-Jagannathan distance (HJ distance hereafter), which is explored in Hansen and Jagannathan (1997), measures a distance between the minimum-norm SDF derived from basis assets and a theoretically specified (and accordingly parametric) SDF.

second step, we estimate the corresponding measure of each individual cryptoasset and evaluate 'where' the estimate is located in the cross-sectional distribution computed in the first stage. This enables us to compute the p-value of the estimated measure corresponding to that particular cryptoasset. If the p-value is greater than a pre-specified, for example, 2.5%, the cryptoasset is classified as domesticated at 97.5% confidence level and, if not, it is classified as non-domesticated.

The estimation results of this novel two-step procedure for distinguishing between domesticated cryptos and non-domesticated ones is used in the following three tests for a horse race competition among the candidate measures. Firstly, we investigate whether the proportion of cryptoassets sorted as domesticated by a certain candidate measure has increased after the above-mentioned structural break. In this test, as expected, the original distance measures, $\frac{\alpha_c^*}{\|\omega_c\|}$ and $\frac{\alpha_c^*}{\|e_c\|}$, result in a decrease rather than an increase in the proportion of domesticated cryptos after the break, which is incompatible with the result of the structural break on comovment. In contrast, the candidate measures adjusted by the linear or quadratic amplification factor, α_c^*, $\alpha_c^*\|\omega_c\|$ and $\alpha_c^*\|e_c\|$, show a sharp increase in the proportion of domesticated cryptoassets after the break. Secondly, we confine the analysis to the post-break period and examine whether the SDF built upon the cryptoassets classified as domesticated by each measure is 'significantly' correlated with m_x^*, the SDF of the original stock basis assets and concomitantly whether the SDF built upon the cryptoassets classified as non-domesticated by that measure is 'trivially' correlated with m_x^*. To estimate those crypto-based SDFs, we need to construct a set of crypto basis assets. We first construct the crypto basis assets by using the cluster analysis suggested by Ahn, Jennifer and Dittmar (2009). In addition, we conduct a bootstrapping analysis wherein the crypto basis assets are randomly selected. A nice feature of the latter analysis is that it enables us to construct the bootstrapped distribution of correlations by which we can test the statistical significance of correlation. Both tests add up to a conclusion that α_c^* is the best athlete as a domestication measure. Finally, to confirm this conclusion once more, we examine the regression of the SDF of the cryptos designated as domesticated by α_c^* against a set of stock market factors (such as the five factors of Fama and French (2015) and the q^5 factors of Hou et al. (2021)) and/or a set of crypto market factors proposed by Liu, Tsyvinski and Wu (2022) and do the similar analysis for the SDF of non-domesticated cryptoassets in a separate regression. The overall results show that the SDF of the domesticated cryptos are strongly related with the stock market factors but not with the crypto-market factors. In contrast, the the estimation results with the SDF of the non-domesticated cryptos are diametrically opposite; the SDF of the non-domesticated cryptos are almost entirely explicable by the crypto-specific factors, but not by the stock market factors at all.

All of the three tests for horse racing deliver the same conclusion: α_c^* is the best-performing measure of domestication. Going back to the issue on the post-break change in the proportion of domesticated cryptos based on α_c^* as a valid measure, we can also make a conclusion that the proportion of domesticated cryptoassets has skyrocketed after the structural break, for example, from 7% to 36% when a conservative domestication criterion, 95% confidence level, is adopted. This upsurge in domesticated cryptos is in sync with the institutionalization documented in the survey literature.

The remainder of this paper is organized as follows. In section 2, we make a brief literature view on the recent studies on cryptoassets and discuss how this study is related to them. Section 3 investigates the econometric state-space model to identify when the cryptoassets spearheaded by Bitcoin begin to show comovement with stocks, the first property of domestication. Therein, we also examine the spill-over effect of the overnight returns of Bitcoin (and MVDA Index) on the S&P500 index and the Nasdaq. In Section 4, we devise a set of potentially admissible candidate measures of domestication based on the two properties characterizing domestication: comovement with stocks and subjection to the pricing rule prevailing in the stock market. The tests for the horse race competitions and their implication are discussed in Section 5. We make concluding remarks in Section 6. The proofs of all lemmas, propositions and corollaries are provided in Appendix.

2 Literature Review

Given a short history of the cryptoasset market, it is only recently that this market has been studied in academia. Broadly speaking, the existing studies center upon two issues: an inter-market relation between the cryptoasset market and other asset markets and an intra-market analysis inside the cryptoasset market. Borri (2019) documents that the major individual cryptoassets are not exposed to the tail risk of other asset classes such as the U.S. equitiy market and gold, while being highly exposed to the tail risk unique to the cryptoasset market. Extending this work, Borri and Santucci de Magistris (2021) find that crypto premium, which is defined as compensation for skewness and kurtosis in the stochastic discount factor is not associated with a set of non-crypto factors such as Fama-French stock market factors, VIX (volatility index) and gold index. Most comprehensive analysis is done in Liu and Tsyvinski (2021), which construct an index of cryptoassets and analyze its risk-return trade-off. Using the time-series analysis, they document that the cryptoassets have no exposure to stock market and macroeconomic factors, not to mention major currencies and commodities. Liu, Tsyvinski and Wu (2022) consider a three-factor model composed of cryptoasset market, size and momentum factors, which is isomorphic to the Cahart's four-factor pricing model except for the book-to-market factor. Their empirical results strongly support for the three-factor model in explaining the 10 long-short strategies based on cryptoasset characteristics known to generate sizable and statistically significant excess returns. In contrast, stock market factors show a limited explanatory power over the same set of strategies. Similarly, Bianchi and Babiak (2021) adopt an instrumented principal component analysis (IPCA) to estimate latent factors which govern a cross-sectional dispersion in risk premia of cryptoassets on a daily frequency. They find that these latent factors are strongly associated with liquidity, size, reversal and downside risks of the cryptoassets, but not related with stock market factors at all.

In sum, the above studies focus on identifying factors underlying the cryptoasset market, either pre-specified factor-mimicking portfolios or latent factors, and examine whether these factors are concatenated with the well-known stock market factors. All of these studies did not find any evidence on such a link and conclude that the crypto-market is segmented from the stock market. In addition, they also document that the factors underlying the cryptoasset market are more deeply related with the previous return behaviors such as momentum, reversal and downside risk.

Our paper differentiates itself from these existing studies in two respects. First, the cryptoasset market is still in an initiatory stage. As a result, this market has been experiencing a continual vicissitude of circumstances, thereby rendering it difficult to assume a stable market wherein the moments of the unconditional joint distributions of the cryptoasset returns do not change when shifted in time. For example, some cryptoassets may be successful in going mainstream on the back of institutionalization and, as a result, they have experienced a structural change in their intra-market relations with other cryptoassets and inter-market relations with other asset classes. As shown in Section 3 and 4, we find that there is, in fact, strong evidence on a structural break at the end of February 2020. Therefore, there is a strong possibility of dramatic changes in their return behavior and their relations with other assets before and after the break. The above-mentioned empirical studies are based on either the pre-break data or the data encompassing the data over both pre- and post-break periods. In contrast, we decompose the data from 2017 toi 2022 into two subperiods, before and after the break. By doing so, we try to highlight the impact of institutionalization on the cryptoassets' inter-market relations with stocks.

Second and related, given its current stage of institutionalization, the cryptoasset market is still in the process of establishing its own identity, thereby too premature to diagnose whether it is fully integrated with other assets classes. Any close association of the cryptoasset with other asset markets is more likely to occur on an individual basis rather than on a collective basis. That is the reason this paper focuses on 'domestication' in lieu of market integration. Consistent with this conjecture, we find that the majority of cryptoassets are yet to be domesticated and thus the overall crypto market itself is still disintegrated with the stock market, which reaffirms the empirical findings of the existing studies. However, on an individual level, some cryptoassets begin to demonstrate domestication and not only move in tandem with stocks but also subject themselves to the pricing rule governing the stock market. Of course, if the majority of the cryptoassets are domesticated, then the cryptoassets would be integrated with the stocks on a market level, and as discussed above, it is still far from integration. In that sense, our study can be thought of as a moving image that empirically documents the ongoing 'evolution' of integration process. In comparison, the existing studies are equivalent to a still image, which empirically documents the evidence of disintegration. Accordingly, our paper is a complementary work to better understand the empirical results documented in the extant literature.

3 Empirical Analysis of Comovment in Price

As discussed in Introduction, a few survey studies document the ongoing institutionalization of the cryptomarket. Then a natural question is whether the institutionalization is strong enough to induce non-fundamentally driven comovement in price between the cryptoassets and traditional assets as suggested by the category view of Barberis, Shleifer and Wurgler (2005). Suppose that for a set of large institutional investors, the cryptoassets belong to the same category in which certain traditional assets are affiliated with. If these investors move money in and out of that category as a whole, the demand pressure makes the prices of the cryptoassets and the traditional assets move in tandem simply due to the bundle trades. As such, the price comovement between Bitcoin and the traditional assets, which are not fundamentally disconnected can be counted as empirical evidence in support of the category view.

Since the institutionalization of the cryptoassets is a recent episode, we investigate whether the cryptoasset market experienced a structural break in terms of its price comovment with traditional assets where the structural break date is unknown. In addition, we examine whether the returns on the cryptoassets affect those on the stocks through a spill-over effect while the stock market is dormant during overnights and weekends. If such a spill-over effect is detected, it is another strong piece of evidence to support the category view. We examine these tests with Bitcoin and the MVDA as the central figures.

3.1 A Structural Break Analysis with an Unknown Break Date

The traditional assets we consider are stocks represented by S&P500 and Nasdaq Composite Index (Nasdaq from here on), gold and dollar index. S&P500 and Nasdaq are selected to reflect the investor sentiment in risky assets. Gold is considered because Bitcoin (BTC) is coined as a digital gold among some proponents based on an argument that there is a finite amount of both and they act as a store of value outside of traditional monetary systems. The dollar index is included given the fact that Bitcoin is initially proposed as a replacement for fiat currencies.[13] We denote S&P500 return or Nasdaque return at time t by S_t. Gold and dollar index returns are referred to as A_t and D_t, respectively. Finally, B_t denotes the Bitcoin return or the MVDA return at time t.

One of the most standard approaches for measuring the comovement among various asset returns is to decompose variance through a dynamic common factor (DCF) model estimation. In our study, we construct a DCF model of (S_t, A_t, D_t, B_t), in which the Bitcoin return dynamics are determined by one global factor, three local factors, and one Bitcoin-specific factor. The global factor is a common factor across all asset returns, whereas a local factor is a local common factor between Bitcoin and a specific category of other assets. All factors are assumed to be mutually independent. The proportion of the total return variance of Bitcoin accounted for by the factors quantifies the comovement between the Bitcoin and financial markets.

The DCF model we estimate is given by

$$\begin{aligned}\begin{pmatrix} S_t \\ A_t \\ D_t \\ B_t \end{pmatrix} &= \begin{pmatrix} G_t \times \gamma_{S,M_t} \\ G_t \times \gamma_{A,M_t} \\ G_t \times \gamma_{D,M_t} \\ G_t \times \gamma_{B,M_t} \end{pmatrix} + \begin{pmatrix} L_t^S \times \lambda_{S,M_t} \\ 0 \\ 0 \\ L_t^S \times \lambda_{B,M_t} \end{pmatrix} + \begin{pmatrix} 0 \\ L_t^A \times \delta_{A,M_t} \\ 0 \\ L_t^A \times \delta_{B,M_t} \end{pmatrix} \\ &+ \begin{pmatrix} 0 \\ 0 \\ L_t^D \times \kappa_{D,M_t} \\ L_t^D \times \kappa_{B,M_t} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ L_t^B \times \psi_{B,M_t} \end{pmatrix} + \begin{pmatrix} E_t^S \\ E_t^A \\ E_t^D \\ E_t^B \end{pmatrix}, \end{aligned} \tag{1}$$

where $\begin{pmatrix} E_t^S & E_t^A & E_t^D & E_t^B \end{pmatrix}'$ follows a multivariate normal distribution, $\mathcal{N}\left(0, \begin{bmatrix} \Sigma_{M_t} & 0 \\ 0 & \sigma^2_{B,M_t} \end{bmatrix}\right)$, and M_t is a first-order non-recurrent two-state Markov process, taking either 0 or 1.[14] The initial state is given by $M_0 = 0$. The transition probability of the state $p_{00} = \Pr[M_t = 0 | M_{t-1} = 0]$ from state 0 to state 0 is to be estimated, whereas $p_{11} = \Pr[M_t = 1 | M_{t-1} = 1]$ is fixed to one. As a result, a regime shift is permanent, and the point in time which the state changes from 0 to 1 is the structural breakpoint. G_t is the global common factor, L_t^S is the local common factor between the Bitcoin and aggregate stock indices (stock-BTC factor), L_t^A is the local common factor between the Bitcoin and gold (gold-BTC factor), and L_t^D is the local common factor between the Bitcoin and dollar index (dollar-BTC factor). L_t^B is the Bitcoin-specific factor, which is unique to Bitcoin. These common factors are assumed to follow mutually independent and serially uncorrelated

[13] In June 2021, El Salvador announced the 'Bitcoin Law,' which made it the first nation to adopt Bitcoin as legal tender and required businesses to accept it as payment.

[14] We also estimated a three-state model in order to examine the possibility of two changepoints, and found that only two states are identified. These two states are the same as those estimated from the two-state model in terms of timing. Thus, we can say that the model with one changepoint is more supported by the data than the model with two changepoints.

normal distributions. That is,

$$\begin{bmatrix} G_t \\ L_t^S \\ L_t^A \\ L_t^D \\ L_t^B \end{bmatrix} \Big| \sigma_G^2, \sigma_S^2, \sigma_A^2, \sigma_D^2, \sigma_B^2 \overset{iid}{\sim} \mathcal{N}\left(0,\ diag(\sigma_G^2, \sigma_S^2, \sigma_A^2, \sigma_D^2)\right). \tag{2}$$

The factor loadings are state-dependent, so that the variance decomposition of the Bitcoin return is subject to regime shifts. The first elements of $\gamma_{S,M_t=0}$, $\lambda_{S,M_t=0}$, $\delta_{A,M_t=0}$ and $\kappa_{D,M_t=0}$ are constrained to be one for identifying the factors and states. L_t^B is the Bitcoin-specific factor and σ^2_{B,M_t} is its conditional variance given the state. Note that because the 3×3 variance-covariance matrix Σ_{M_t} is non-diagonal, the vector of $(E_t^S,\ E_t^A,\ E_t^D)$ simultaneously captures the asset-specific factors and the common factors between $(S_t,\ A_t,\ D_t)$ other than the four common factors. Intuitively, if $\gamma_{B,M_t=0} = \lambda_{B,M_t=0} = \delta_{B,M_t=0} = \kappa_{B,M_t=0} = 0$, it implies no price comovement of the Bitcoin with other financial markets before the break. In contrast, strongly non-zero $\gamma_{B,M_t=1}$, $\lambda_{B,M_t=1}$, $\delta_{B,M_t=1}$, or $\kappa_{B,M_t=1}$ indicates the price comovment of the Bitcoin with other financial markets after the break, which is a necessary condition for domestication of the Bitcoin.

Our econometric approach is Bayesian, and we complete our Bayesian modeling by specifying priors of the model parameters. Basically, our priors of the model parameters are weak and symmetric between the states, so that the regime changes are identified and detected by the information in the data rather than prior. The prior of all factor loadings is $\mathcal{N}(0,\ V_\beta = 1)$ and the prior of Σ_{M_t} is an inverse-Wishart distribution, $\mathrm{IW}(R_0 = 2 \times I_2,\ v_0 = 2)$ for all states. The transition probability is assumed to follow a beta distribution, $Beta(a_0 = 2000, b_0 = 2)$, based on our prior belief that the first state is highly persistent. Finally, the variances of the common factors and Bitcoin-specific factor are assumed to follow an inverse-gamma distribution, $IG(\alpha_0 = 2, \delta_0 = 2)$.

For posterior inference, we run 12,000 MCMC iterations, and use 10,000 posterior samples after discarding the first 2,000 samples. The sample period is from January 20, 2015 to December 31, 2021. The posterior sampling algorithm is provided in Appendix B. The maximum inefficiency factor of the model parameters is less than 5, so the effective simulation size is larger than 2,000. This indicates convergence and good mixing of the MCMC chain.

Figure 1 shows the posterior probability estimate of the second state (i.e., state 1) over time. As this figure displays, the posterior mode of the changepoint is estimated at February 21st of 2020. To visualize the occurrence of the structural break, we also plot the time-varying correlations between the Bitcoin and S&P500 returns with the moving window of 100 business days. Before the break, the rolling correlations used to fluctuate around zero, whereas they never fall below zero after the break. In addition, the posterior probability coupled with the rolloing correlation shows that the structural break took place all at once rather than gradually; the domestication of Bitcoin, if it occurs, seems to be consummated at a tipping point at which a sequence of small changes becomes significant enough to cause a structural change. Figure 2 summarizes major critical events that have befallen the crypto market. The derivatives exchanges, CBOE and CME, moved first with the futures on Bitcoin, which is followed by a series of participation of major financial institutions in the market. We can say that the cumulative pressure of institutionalization along the continuum ultimately triggers the structural break in Figure 1.

Then, the following question is the source of the structural break. To see this, we present the variance decomposition result of Bitcoin return in Table 1, computed by the posterior samples of the model parameters. The table shows that the increased comovement of Bitcoin return with other asset returns a driving nature of the structural break. Before the break, the Bitcoin-specific factor explains the vast majority, 95.6% of Bitcoin return fluctuations. The other assets such as stock, gold and dollar do not have any noticeable relation with Bitcoin. However, after the break, the stock-BTC factor accounts for 62.7% of the Bitcoin return variation whereas he role of the Bitcoin-specific factor plunges to 24.9%. This results shows that the structural break took place mainly on the back of strong positive comovment beween Bitcoin and S&P500.

The structural break date is critically important in the following analyses, so that we re-estimate it with the MVDA index, a market-cap weighted index composed of top 100 cryptoassets distributed by the MVIS CryptoCompare. The estimation results are quite similar, albeit slightly different in the role of the gold-MVDA factor. The estimated break date is still the same, February 21st of 2020. Again, the traditional assets were lack of any explanatory power before the break. In contrast, herein not only the S&P500-MVDA factor but also the gold-MVDA factor is significant in its explanatory power. In sum, February 21st of 2020 seems to be a convicing estimate of the timing of a structural change and such a structural break took place on the back of comovement between cryptoassets and stocks, above all things.

3.2 Overnight/Weekend Return Analysis

Given the above result, we narrow down our focus to the price movement of Bitcoin in tandem with stocks represented by the S&P500 and Nasdaq Indices. In particular, we utilize a unique feature of the cryptoassets that they trade around the clock across the globe unlike stocks. Specifically, we explore whether the overnight returns on Bitcoin predicts the overnight returns on S&P500 and Nasdaq from the closing price of the previous trading day to the opening price on the following business day during which stocks are closed. If such a relationship is detected, the spill-over effect between Bitcoin and stocks may not be entirely unidirectional, from stocks to cryptoassets, but partially reciprocal; the crypto market, which is always awake, may carry the prevailing market sentiment on behalf of the 'category' of risky assets while the stock market is asleep. This could be decisive evidence that the investor sentiment is a common factor co-shared by the stocks and the cryptoassets.

Stock markets are dormant over longer periods of time, either on a regular basis or on an irregular basis: weekend days and holidays. These weekend/holiday returns are already included as a part of the above-mentioned overnight return data. However, it would be interesting to see whether the spill-over effect is still at work over such longer periods of time. As such, we single out the weekend and holiday returns out of the overnight returns and do the same analysis separately.

In testing this relation, we include the previous day's intraday returns on S&500 or Nasdaq to control for a potential of autocorrelation in the stock indices. In addition, given the fact that the U.K.'s stock market opens well before the U.S. market, we further include the part of the intraday return on the FTSE100 index to control for the spill-over effect of the U.K.'s stock market on the U.S. market.

Specifically, Figure 3 illustrates the time periods over which the returns on the relevant assets are measured. S_t^{US}, the overnight return on the stock index (S&P500 or Nasdaq), is computed for the period from the closing of the New York Stock Exchange (NYSE) on the previous day (EST

16:00) to the opening of the NYSE on the next business day (EST 09:30). Similarly, R_t^B is the return on Bitcoin, the holding period of which overlaps with S_t^{US}. Given that S_t^{US} is regressed against R_t^B (along with other explanatory variables), the regression result will tell us whether the price movement of Bitcoin is able to account for the opening price of the U.S. stocks. We also compute its corresponding realized volatility, RV_t^B, which is computed from return observations with frequencies of 1 minute.

As mentioned above, we compute S_t^{UK}, the return on the FTSE 100 index between the opening of the London Stock Exchange (EST 03:00) and the opening of the NYSE (EST 09:00), to net out the spill-over effect of the U.K. stock market on the U.S. stock market. We also compute its corresponding realized volatility, RV_t^{UK}, which is also computed from the return data with 1 minute frequency. Finally, to control for a potential of autocorrelation in the stock index, we compute S_{t-1}^{US}, which is the intraday return on the stock index on the previous day (i.e., the return from EST 09:30 to EST 16:00 on the previous day).

Based on these data, we estimate the following autoregressive distributed lag model with stochastic volatility (ADL-SV model). $\mathcal{Y}_t$ is the observations available right at the opening of the NYSE. θ denotes the model parameters. Then, the data generating process is given by

$$\begin{aligned} S_t^{US} &= \mu_r + S_{t-1}^{US}\rho + R_t^B\left(\alpha_0^B + \mathbb{1}_t\alpha_1^B\right) + S_t^{UK}\alpha_0^{UK} + \exp(h_t/2)\varepsilon_t \\ h_t &= \mu_h + h_{t-1}\phi + \log RV_t^B\left(\beta_0^B + \mathbb{1}_t\beta_1^B\right) + \log RV_t^{UK}\beta_0^{UK} + e_t \end{aligned} \tag{3}$$

where

$$\left(\begin{array}{c} \varepsilon_t \\ e_t \end{array}\right) \Bigg| \mathcal{Y}_t, h_{t-1}, \theta \sim \mathcal{N}\left(0, \left[\begin{array}{cc} 1 & 0 \\ 0 & \sigma^2 \end{array}\right]\right), \text{ and } \mathbb{1}_t = \begin{cases} 0 & \text{if } t = \text{before the break} \\ 1 & \text{if } t = \text{after the break} \end{cases}$$

where h_t is the log stochastic volatility at time t. We introduce a dummy variable $\mathbb{1}_t$, in order to examine whether there exits a structural change in the spill-over effect after the break where we take February 21, 2020 as the breakpoint following the estimation result in the above structural break analysis.

We assume a weak prior of the model parameters such that

$$\begin{gathered} \mu_r,\ \rho,\ \alpha_0^B,\ \alpha_1^B,\ \alpha_0^{UK},\ \beta_0^B,\ \beta_1^B,\ \beta_0^{UK},\ \mu_h \sim \mathcal{N}(0, 0.25) \\ \phi \sim \mathcal{N}(0.95, 0.25), \text{ and } \sigma^2 \sim IG(0.5, 0.1), \end{gathered}$$

where $IG(\cdot,\cdot)$ is an inverse gamma distribution characterized by two parameters: a shape parmaeter and a scale parameter. All the parameters are mutually independent *a priori*. The prior mean of ϕ is 0.95, considering the persistence of volatility that financial asset returns commonly exhibit.

We consider two aggregate stock returns: S&P500 and Nasdaq. The sample period is the same as in the above structural break analysis, but the sampling frequency is daily instead of weekly. We estimate the model parameters and stochastic volatility using the posterior sampling algorithm suggested by Omori et al. (2007).

The estimation results are presented in Table 2. The key parameter we are interested in is $\alpha_0^B + \alpha_1^B$, which detects the explanatory power of the Bitcoin returns on the overnight returns on the U.S. stocks after the break. The key finding is that the Bitcoin return has a significant explanatory power of the U.S. overnight stock returns after the break regardless of whether the U.K. stock market effect is controlled or not. The coefficient of R_t^B in the mean equation has a positive posterior mean and small posterior standard deviation. Specifically, as for the S&P500, the coefficient shows a dramatic increase from -0.001 to 0.020 in the absence of the FTSE100 and from 0.000 to 0.011 in the presence of the FTSE100 as a controlling variable. The post-break coefficients are so precisely estimated that the 99% credibility intervals do not include zero, whereas the pre-break ones are all insignificant. Although the scales of the R_t^B coefficient look much smaller than those of S_t^{UK}, they are not comparable on a par, considering that the fluctuation of Bitcoin return is much larger than the stock return. One thing to mention is that the spill-over effect of Bitcoin on S&P500 is demonstrated in the volatility before the break but such an effect vanished after the break. That is, Bitcoin changed its sphere of influence from the volatility to the mean of the S&P500 overnight returns in the wake of the structural break.

The estimation results on Nasdaq are similar but the scales of the spill-over effect are somewhat stronger as shown in the bottom panel of Table 2. The posterior means jumped from -0.001 to 0.045 in the absence of the FTSE100 and from 0.000 to 0.032 in the presence of the FTSE100. As such, the post-break coefficients are larger than those in the case of S&P500. All of them are statistically significant whereas the pre-break means are not, which is consistent with the estimation results in the case of S&P500. Overall, we can summarize that the spill-over effect of Bitcoin is more evident in Nasdaq, which is heavily prone toward the technology sector as its constituents.

Table 3 summarizes the estimation results on the weekend/holiday returns. The overall results are similar to those in Table 2; after the break, the spill-over effect of Bitcoin on the U.S. stock indices are alive and well over the longer periods of time when the U.S. stock markets are closed, whereas they are not before the break. In the upper panel, the post-break sensitivity of the S&P500 to Bitcoin, which is captured by R_t^B, is 0.020 and 0.010, without and with being controlled for the FTSE100 respectively. These estimates are similar to their corresponding values on the overnight returns, 0.020 and 0.011 presented in Table 2. As such, the spill-over effect seems to be robust to the time span of dormant stock market. However, this conclusion is overturned in the estimation results on the Nasdaq, which is tabulated at the bottom panel of Table 3. The post-break coefficients of R_t^B are 0.057 and 0.035 without and with the FTSE100 respectively, and these estimates are larger than the corresponding estimates in the analysis on the overnight returns, 0.045 and 0.032 presented in Table 2. Therefore, the spill-over effect is rather strengthened, if any, over the longer horizon in the Bitcoin-Nasdaq relationship. This result insinuates the fact that the crypto market represented by Bitcoin is more likely to move in tandem with technology stocks, which share the market sentiments with it, in response to fresh developments in the economy.

4 Measurement of Domestication: Theory

A full-scale analysis of domestication traces its history back to the study of Darwin (1875), who tried to identify a suite of altered morphological, behavioral and physiological traits that are common across domesticated animals. He referred the cluster of such altered traits exhibited by domesticated species as 'domestication syndrome.' Herein, we also attempt to develop an admissible measure of domestication by identifying the traits that are shared by domesticated assets, but not by non-domesticated ones. Then, what would be the generic traits that should be displayed by domesticated assets in the financial market? Undoubtedly comovement could be counted as the most important trait required for domestication. If the price of a cryptoasset fluctuates independently of the domesticated market, we can say that it trades in a disparate and segmented market of its own and thus is not tamed yet. The structural break analysis coupled with the ADL-SV estimation in the previous section indicates that Bitcoin has begun to co-move in price with the major stock indices since the structural break it experienced at the end of February, 2020. Therefore, we can say that Bitcoin currently exhibits this trait. However, comovement may not be a unique trait that should be carried by domesticated assets. Once comovement is accepted as a trait, it also invites a concomitant trait, the subjection of the cryptoasset to the price system prevailing in the domesticated market. If the cryptoasset co-vary in price with the domesticated assets, its price would not be determined in isolation any more and should be determined via the covariances with the domesticated assets, which are not trivial anymore. That is, investors evaluate its price by referring to the value of the domesticated assets as benchmarks where the prices of those domesticated assets are governed by a certain pricing rule.

If we accept these two traits as the domestication syndrome, a domestication measure, which distinguishes between domesticated cryptoassets and non-domesticated ones, should satisfy the two traits. To develop such a measure, we begin with the premise that the stock market is incomplete and is represented by a set of stock basis assets, each constituent of which is assumed to be 'already domesticated.' In addition, we assume that the price system in such a market is viable, i.e., not allowing for any violation of the Law of One Price (LOP hereafter), or equivalently guaranteeing the existence of the weak-form SDFs which correctly price all the stock basis assets. Then we release an individual cryptoasset into the stock market and price it using the minimum-norm SDF among the above-mentioned SDFs. Theoretically speaking, the resulting pricing error associated with the cryptoasset does not necessarily mean that the crypto is genuinely mispriced. Unless its payoff is perfectly replicable by the stock basis assets, we can always find another SDFs which can correctly price not only the basis assets but also the given cryptoasset. That is, the admissible set of SDFs shrinks when the set of basis assets is augmented by the cryptoasset in consideration. Then, in light of the series of seminal studies contributed by Hansen and Jagannathan, a natural candidate for a domestication measure would be a distance measure, which measures the degree of contraction of the admissible set of SDFs. The resulting distance measure is akin to the amount of upward shift in the HJ bound or the HJ distance measure. Then a remaining question is whether this distance measure satisfies the above-mentioned two traits. This is what we will analyze in this section. For brevity's sake, we use 'cryptoasset' and 'crypto' interchangeably.

4.1 Market Environment

We consider an economy characterized by a probability space triplet $(\Omega, \mathbb{F}, \mathscr{P})$ where $\mathbb{F}$ is a sigma algebra of subsets of Ω, and $\mathscr{P}$ is a probaility measure on $(\Omega, \mathbb{F})$. We also define $\mathscr{L}^2$, the linear space of square-integrable random variables on the triplet. $\mathscr{L}^2$ is endowed with its usual inner product $\langle x_1 | x_2 \rangle = E(x_1 x_2)$ and norm $\|x\| = \sqrt{E(x^2)}$ and thus $\mathscr{L}^2$ is a Hilbert space.

Assumption 1: *There are N 'domesticated' basis assets, the payoffs of which are denoted by $\{x_i\}_{i=1}^N$. $x_i \in \mathscr{L}^2 \ \forall \ i$ is a total return on a dollar investment in basis asset i. The basis assets do not include any redundant assets, i.e., $E(xx')$ has a full rank, where $x = (x_1 \ x_2, \cdots, x_N)'$.*

Assumption 1 states that each of these N basis assets is fully domesticated. Without loss of generality, we assume that these assets are stocks, thereby calling them 'stock basis assets.' As such, a crypto, if domesticated, should display the similar traits that are shared among stocks.

Based on these basis assets, we define the following *attainable set*, $\mathscr{A}(x)$, which is a payoff set of all portfolios constructable based on the basis assets, x:

$$\mathscr{A}(x) = \left\{ y \in \mathscr{L}^2 \text{ such that } y = \theta' x, \;\; \theta \in \mathbb{R}^N \right\}.$$

The attainable set is conditional upon the basis assets simply because the set of obtainable portfolios are determined by the basis assets. If $\theta' \ell_N = 1$ where 1_N is a N-dimensional column vector of ones , θ becomes a vector of portfolio weights. Since it is a set of payoffs, $\mathscr{A}(x) \subset \mathscr{L}^2$. In particular, set $\mathscr{A}$ increases with the number of basis assets, i.e., $\mathscr{A}(x_N) \subset \mathscr{A}(x_M)$ if $x'_M = (x'_N \;\; x'_{M-N})$ and $\mathscr{A}(x_{M-N}) \not\subset \mathscr{A}(x_N)$ where $M > N$. Simply put, adding more assets to the existing basis assets enlarges the attainable set as long as the additional assets are not *perfectly* replicable by the exisiting basis assets. We assume that even when we include all available marketed assets, the market is still incomplete:

Assumption 2: *The capital market is incomeplete even when all traded assets are adopted as basis assets such that*

$$\max_x \mathscr{A}(x) \subset \mathscr{L}^2 \text{ and } \max_x \mathscr{A}(x) \neq \mathscr{L}^2.$$

This assumption will be shown to be critical below. We also define the price functional associated with $\mathscr{A}(x)$ such that

$$q(y) = \left\{ \theta' \ell_N \text{ such that } y = \theta' x \text{ and } \theta \in \mathbb{R}^N \right\}.$$

The price functional indicates that all element of $\mathscr{A}(x)$ is marketable, i.e., its fair value must be a singleton, otherwise it violates the LOP, a necessary condition for a stricter no-arbitrage condition. Following the seminal work of Hansen and Jagannathan (1991), we establish the following lemma on the existence of the global minimum-norm (GMN) SDF on the basis assets:

Lemma 1: *Given the mutually non-dependent domesticated basis assets, x, there always exists a unique attainable stochastic discount factor, $m_x^* \in \mathscr{A}(x)$ which satisfies*

$$E(m_x^* \cdot x) = \ell_N,$$

and $m_x^* = \theta_x' x$ where $\theta_x = E(xx')^{-1} \ell_N$. In addition, the set of admissible m_xs, which satisfy $E(m_x \cdot x) = \ell_N$ is a set of orthogonal extensions of m_x^* such that

$$\mathscr{M}(x) = \left\{ m_x | m_x = m_x^* + \varepsilon_x \textit{ where } \varepsilon_x \in \mathscr{A}(x)^{\perp}, \right\}$$

where $\mathscr{A}(x)^{\perp}$ is an orthogonal complement of $\mathscr{A}(x)$. Since the market is incomplete, $\mathscr{M}(x)$ is not a singleton set, i.e., $\mathscr{M}(x) - \{m_x^*\} \neq \emptyset$. m_x^* is the minimum-norm SDF, $\|m_x^*\| = \min \|m_x\|$.

Strictly speaking, not all of $m_x \in \mathscr{M}(x)$ are valid SDFs. Harrison and Kreps (1979) show that the existence of m_x rules out any violation of the LOP condition, but does not rule out an opportunity of arbitrage transactions. The no-arbitrage condition states that nonnegative payoffs which are strictively positive with positive probability should be positively priced. To gaurantee the absence of such arbitrage opportunity, an additional restriction is needed such that the SDF should be positive with probability one, i.e., $m_x \gg 0$. The set of strictly positive SDFs is defined as $\mathscr{M}_{++}(x) = \{m_x | E(m_x \cdot x) = \ell_N \text{ and } m_x \gg 0\}$ and thus $\mathscr{M}_{++}(x) \subset \mathscr{M}(x)$. In that sense, $m_x \in \mathscr{M}(x)$ is called a weak-form SDF while $m_x \in \mathscr{M}_{++}(x)$ is called a strong-form SDF. In this paper, we focus on the weak-form SDF for econometric tractability that will be discussed in the next section.

4.2 Domestication Syndrome: Two Traits

Into the above financial market, we introduce a particular *individual* crypto, c, which is characterized by its payoff, $x_c \in \mathscr{L}^2$. We make the following assumption for practical convenience.

Assumption 3: x_c is not attainable by the basis assets, i.e., $x_c \notin \mathscr{A}(x)$.

Thus, we assume that the crypto's payoff is not perfectly replicable by the basis assets. From the projection theorem, x_c can be decomposed into

$$x_c = P_x(x_c) + \omega_c,$$

where $P_x(\cdot) : \mathscr{L}^2 \to \mathscr{A}(x)$ is the projection operator from any point in $\mathscr{L}^2$ onto $\mathscr{A}(x)$. $P_x(x_c) \in \mathscr{A}(x)$, and therefore $\omega_c \in \mathscr{A}(x)^{\perp}$. Assumption 3 is equivalent to a statement that $\omega_c \neq 0$. Because $P_x(x_c) \in \mathscr{A}(x)$, there exists some $\phi_c \in \mathbb{R}^N$ such that $P_x(x_c) = \phi_c' x_c$ where $\phi_c = E(xx')^{-1}E(xx_c)$.

Let x_a denote an x_c-augmented payoffs, $x_a = (x' \ x_c)'$. Then, $\mathscr{A}(x_a)$ and $\mathscr{M}(x_a)$ refer to its corresponding attainable set and the set of weak-form SDFs respectively. Assumption 3 dictates the following lemma:

Lemma 2: *$\mathscr{A}(x) \subset \mathscr{A}(x_a)$ and $\mathscr{A}(x) \neq \mathscr{A}(x_a)$ under Assumption 3.*

Lemma 2 is a trivial outcome of the property that if any non-dependent asset payoff is added, the extended attainable set is enlarged in a strict sense.

Lemma 3: *$\mathscr{M}(x_a) \neq \emptyset$ under Assumption 3. In addition, $\mathscr{M}(x) \supset \mathscr{M}(x_a)$ and $\mathscr{M}(x) \neq \mathscr{M}(x_a)$.*

Lemma 3 states that we can always find a set of admissible SDFs that can correctly price the x_c-augmented basis assets; that is, they can price not only the stock basis assets but also and the given crypto. In addition, because such SDFs are required to price the crypto on top of the stock basis assets, more restrictions are imposed on the SDFs, thereby the resulting admissible set of SDFs being smaller.

Now that we understand a change in the admissible set of SDFs induced by the individual crypto, we are ready to discuss the economically legitimate measure of domestication. Intuitively, a non-domesticated cryptoasset is isolated and accordingly its price movements are not related to the stock basis assets. That is, the non-domesticated cryptoasset, like a wolf, maintains its own indigenous 'wildness.' In contrast, if the cryptoasset becomes sufficiently domesticated, then it should show a domestication syndrome, which refers to a suite of certain traits as discussed above. We suggest two critical traits that should be shared by domesticated cryptos: comovment and price alignment. We discuss these two traits below.

(1) First Trait: Comovement

The process of domestication involves adaptation. It begins with capturing and keeping the wild animals in captivity and making them adapt to a captive environment. Price (1984) states that the transition from free-living to captive status is often accompanied by changes in availability and accessibility of shelter, space, food and water and by changes in predation and the social environment. These changes result in the development of the domestic phenotype, the observable traits of the animals as a consequence of environmental influences. In a similar vein, comovment is a necessary trait displayed by a croptoasset undergoing domestication once it is kept in captivity by a large community of institutional profession.

Specifically, with respect to anmial behavior, domestication has altered certain behaviors because of man's role as a buffer between the animal and its environment. Similarly, in the realm of cryptos, price behavior may be altered because of institutional investors' role as a buffer between the crypto and its initial market environment. A clue could be found in the category view of comovment proposed by Barberis and Shleifer (2003). The institutionalization of a cryptoasset implies that a large community of institutional investors incorporates it as a component asset at par with stocks in their portfolios. Since they trade this category as a 'bundle', they buy and sell the cryptoasset simultaneously with stocks. This trade induces common factors in the returns of the cryptoasset and the stocks. All in all, the price co-movement is a prominent sign that the cryptoasset is sufficiently institutionalized and domesticated by the 'at-scale' participation of institutional investors who allocate funds across at the level of this category. As such, comovement in price is the most indisputable trait that a crypto should display once it is domesticated.

(2) Second Trait: Price Alignment

Strictly speaking, domestication is different from taming. Driscoll, Macdonald and O'Brien (2009) state that taming is conditional behavioral modification of an individual whereas domestication is permanent genetic modification of a bred lineage that leads to a heritable predisposition toward human association. The existing models of animal domestication are separated between domestication as constituting a form of domination or a type of mutualism. However, revisiting the metaphor of the *domus*, which means house in Latin, domesticated animals are, in general, cultivated to live symbiotically alongside humans, thereby being predisposed to be tame. Similarly, cryptoassets, if fully domesticated, should exhibit a trait of being tamed such that their prices are subject to the pricing rule prevailing in the existing domesticated market, i.e., the stock market in our case. Of course, this price alignment of the cryptos to the stock market comes into play only if the first trait is sufficiently manifest.

(3) Intuition Behind Economic Representation of Domestic Syndrome

Here we discuss how to accommodate the above-mentioned two traits of domestication in the existing framework of asset pricing. The purpose is to establish some key economic intuitions, thereby its analysis somewhat heuristic and not rigorous. To do so, we begin with considering a crypto, the price of which is $\mathbb{P}_c$. Its gross return is x_c. It is not fully domesticated, thereby $x_c \notin \mathscr{A}(x)$. However, from Lemma 3 we know that there exists at least one $m_{x_a} = m^*_x + \varepsilon_c \in \mathscr{M}(x_a)$ where ε_c is an orthogonal extension of m^*_x which validates the market prices of the stock basis assets and the crypto. In order to emphasize that such an extension is specific to the particular cryptoasset, x_c, we change its subscript from x in Lemma 1 to 'c.' The fundamental valuation equation of the crypto asset is $\mathbb{P}_c = E[(m^*_x + \varepsilon_c) \cdot \mathbb{P}_c x_c]$, which results in:

$$\begin{aligned} 1 &= E[(m^*_x + \varepsilon_c) \cdot (P_x(x_c) + \omega_c)] \\ &= E[m^*_x \cdot P_x(x_c)] + E[\varepsilon_c \cdot \omega_c]. \end{aligned} \tag{4}$$

Equation (4) states that the normalized price of the crypto, one dollar, is composed of two components. In the first term, $P_x(x_c)$ is the projected component of x_c and belongs to $\mathscr{A}(x)$. In other words, it is the part of the crypto's payoff, which is replicably by the stock basis assets and such a replication, $\phi_c' x$, is built upon its comovement with the basis assets, $\phi_c = E(xx')^{-1}\boxed{E(xx_c)}$. At the same time, this is the part of the payoff which should be subject to the incumbent pricing rule governing the stock market, m_x^*. If so, $E[m_x^* \cdot \phi_c' x] = \phi_c' \ell_N$ and thus the two terms in Equation (4) can be broken down into the following two separate fundamental valuation equations:

$$
\begin{aligned}
\phi_c' \ell_N &= E[m_x^* \cdot P_x(x_c)] && (5)\\
1 - \phi_c' \ell_N &= E[\varepsilon_c \cdot \omega_c]. && (6)
\end{aligned}
$$

This decomposition of the price of x_c is seemingly natural, but it has a profound economic implication. When an investor purchases this crypto, she conceives of it as a portfolio or bundle of two separate assets, $P_{x|c}$ and ω_c. As such, the payment of one dollar is tantamount to a combination of two separate payments, $\phi_c' \ell_N$ for the replicable payoff $P_x(x_c)$ and $1 - \phi_c' \ell_N$ for the non-replicable payoff ω_c. If Equation (5) holds, the pricing alignment is satisfied. In contrast, Equation (6) shows how ω_c, the non-attinable part of x_c, is priced. How to interpret Equation (6)? Equation (6) is not a pricing equation. That is, it does not say that the outcome of solving the right-hand-side of Equation (6) is $1 - \phi_c' \ell_N$. Rather, it is the other way around. If Equation (5) holds, an econometrician is left with $1 - \phi_c' \ell_N$, the unexplained remaining dollar amount, $1 - \phi_c' \ell_N$. Then, ε_c is determined to satisfy Equation (6) and eventually Equation (4). That is, ε_c is not an observable variable but a sort of slack variable, which is determined to equate $1 - \phi_c' \ell_N$ to $E[\varepsilon_c \cdot \omega_c]$ in Equation (6); it plays a role of a buffer between $1 - \phi_c' \ell_N$ and ω_c.

Under the assumption that Equation (5) holds, $1 - \phi_c' \ell_N$ can be regarded as an inverse measure of comovement, albeit indirect. Its underlying reasoning is as follows. Since the sum of the two component assets' prices are equal to a dollar by normalization, $\phi_c' \ell_N$ is the relative weight of the replicable payoff in the price of x_c, whereas $1 - \phi_c' \ell_N$ is the relative weight of the non-replicable payoff in the price of x_c. That is, $\phi_c' \ell_N$ and $1 - \phi_c' \ell_N$ measure how much portions of the crypto's price are determined by the replicable payoff and the non-replicable payoff respectively. To better understand this, suppose the investor pays the entire price of x_c, one dollar, as a compensation for $P_x(x_c)$ and is not willing to pay any for ω_c. In such a case, $1 - \phi_c' \ell_N = 0$. The investor is not willing to pay any for the non-replicable payoff since this risk might be diversified away once this crypto is incorporated into his or her optimal portfolio. That is, if the institutionalization of this particular crypto is fully accomplished, $\phi_c' \ell_N = 1$ and ϕ_c becomes a vector of investment weights on the stock basis assets. In sum, we can say that Equation (5) reflects the price alignment trait and $\phi_c' \ell_N = 1$ is the comovment trait.

The above interpretation of $1 - \phi_c' \ell_N$ as a measure of comovment is valid only if the pricing alignment is satisfied; i.e., Equation (5) holds. Unfortunately, we do not know whether $\phi_c' \ell_N$ is the actual price that the investor pays for $P_x(x_c)$. All we can observe is the one-dollar bill that the investor pays for the entire payoff, x_c. To discern how much he or she actually pays for $P_x(x_c)$, $P_x(x_c)$ should be traded individually, i.e., it should be marketable, but it is not. Therefore, we cannot directly test whether (5) holds or equivalently whether the pricing alignment is satisfied. However, despite the impossibility of testing Equation (5), $1 - \phi_c' \ell_N$ still offers a solid foundation for the two critical domestication traits. To see this, suppose that the investor actually pays the (non-normalized) price for the attaianble payoff, $(\phi_c' \ell_N + \eta_c)\mathbb{P}_c$ for $P_x(x_c)$. That is, in terms of percentage of price, it is mispriced by η_c. To focus on the impact of η_c, we assume that Equation (6) still holds in the sense that the marginal investor who determines the equilibrium price prices ω_c by using ε_c as the pricing kernel. In such a case, an econometrician, who is not aware of the mispricing, projects the crypto's payoff on the stock basis assets and establishes the following valuation equation based on an alternative orthogonal extension, $\varepsilon_{c\eta}$:

$$\begin{aligned}(1+\eta_c)\mathbb{P} &= E[(m_x^* + \varepsilon_{c\eta})(P_c(x_c) + \omega_c)\mathbb{P}] \\ \Longrightarrow 1 &= E\left[m_x^* \cdot \frac{P_x(x_c)}{1+\eta_c}\right] + E\left[\varepsilon_{c\eta} \cdot \frac{\omega_c}{1+\eta_c}\right],\end{aligned}$$

and the econometrician continues to decompose it into

$$\phi'_{c\eta}\ell_N = E\left[m_x^* \cdot \frac{P_x(x_c)}{1+\eta_c}\right] \tag{7}$$

$$1 - \phi'_{c\eta}\ell_N = E\left[\varepsilon_{c\eta} \cdot \frac{\omega_c}{1+\eta_c}\right]. \tag{8}$$

Then, the wrongly estimated projection parameters, $\phi_{c\eta}$ is, in fact, $\phi_c/(1+\eta_c)$, and thus $1-\phi'_{c\eta}\ell_N = 1 - \frac{\phi'_c\ell_N}{1+\eta_c}$. Accordingly, the econometrician becomes to retrieve $\varepsilon_{c\eta}$ from the following equation:

$$E\left[\varepsilon_{c\eta} \cdot \omega_c\right] = 1 - \phi'_c\ell_N + \eta_c, \tag{9}$$

which is equivalent to $\eta_c + E[\varepsilon_c \cdot \omega_c]$. Therefore, in the presence of mispricing in the crypto, $\varepsilon_{c\eta}$ is determined to reflect not only the comovment, $1-\phi'_c\ell_N$ but also the pricing error, η_c. Of course, the econometrician is not able to distinguish these two. What Equation (9) and equivalently Equation (8) tell us is that $1 - \phi'_{c\eta}\ell_N$ (or $\varepsilon_{c\eta}$ that is determined by it) is the key composite variable, which contains information about the both domestication traits, comovement and pricing alignment. Therefore, a desirable measure of domestication should be built upon $\varepsilon_{c\eta}$.

Before moving on, one thing to mention is ω_c. ω_c is the crypto-specific idiosyncratic component of x_c, which reflects the replication error. If the crypto were perfectly replicable by the stock basis assets, it would be zero. In such a case, the crypto would be a perfect target of a genuine arbitrage transaction. If not, the arbitragers may consider a statistical arbitrage, which is exposed to the volatilty of ω_c. However, if its volatility is excessively large, they would not consider this crypto as a target for an arbitrage transaction because of a concern about the excess risk that they have to bear during the transaction. As such, its volatility is a *de facto* measure of comovement of the crypto from the perspective of investment professions; the larger the volatility is, the less strongly the crypto comoves with the stock basis assets. Given that the volatility of $e_{x|c}$ reflects comovement, it is highly like to be positively associated with $|1 - \phi'_c\ell_N|$, which subsumes information regarding comovement along with pricing alignment. If such a positive correlation is sufficiently strong, that is another piece of evidence that $|1 - \phi'_c\ell_N|$ carries information about comovement. Of course, this is an empirically testable question and will be explored in the next section.

4.3 Nonparametric Domestication Measures

Following the above discussion, we construct a set of candidate domestication measures based on the two traits, comovement and price alignment. We *a priori* remove parametric measures from the list of candidates for two reasons. Firstly, a parametric measure requires a theoretical SDF delivered by a particular asset pricing model. However, such an approach is subject to a misspecification error of the adopted asset pricing model, which could be excessively substantial.[15] Secondly, due to the lack of accounting chacteristics associated with the cryptoassets, the popular Fama-French three-factor or five-factor models and the q-factor model are not applicable. Consequently, we have no choice but to rely upon the nonparametric approach in building up a domestication masure. In addition, as discussed in the previous section, ε_c, the orthogonal extension of the SDF, is the key composite variable which retains the information regarding the two domestication traits, comovement and pricing alignment. As such, we center upon ε_c in search of a domestication measure.

(1) Mean-Second Norm Bound of SDFs

We begin with a full-fledged formal analysis of Equation (4) in a more general framework. We extend the global minimum norm SDF, m_x^*, to a full set of SDFs which differ in their means, μ_m.

Lemma 4: *Given the the stock basis assets, x and an arbitrary expected value of the SDF, μ_m, the minimum-norm SDF is*

$$m^*_{x|\mu_m} = \mu_m + (\ell_N - \mu_m \mu_x)' \Sigma_x^{-1} (x - \mu_x),$$

and its corresponding squared norm is

$$E\left[m^{*2}_{x|\mu_m}\right] = \inf \|m_{x|\mu_m}\|^2 = (1 + a_x)\mu_m^2 - 2b_x\mu_m + c_x, \tag{10}$$

where

$$a_x = \mu_x' \Sigma_x^{-1} \mu_x, \quad b_x = \mu_x' \Sigma_x^{-1} \ell_N, \quad \text{and} \quad c_x = \ell_N' \Sigma_x^{-1} \ell_N,$$

and μ_x and Σ_x refer to the mean vector and the covariance matrix of the stock basis assets' payoffs respectively.

Lemma 4 basically describes the HJ bound with slight adjustment. For convenience, we modify the HJ bound from the original mean-standard frontier of the SDF to the corresponding mean-second norm frontier of the SDF. Equation (10) shows that the relation between μ_m and $\|m^*_{x|\mu_m}\|^2$ is parabolic. It can be rewritten as:

$$\frac{E(m^{*2}_{x|\mu_m})}{1 + a_x} - \left(\mu_m - \frac{b_x}{1 + a_x}\right)^2 = \frac{(1 + a_x)c_x - b_x^2}{1 + a_x}. \tag{11}$$

$a_x > 0$ from the positive definiteness of Σ_x^{-1}, which comes from the fact that Σ_x itself is positive definite. Therefore, the mean-second-norm frontier of the SDF, i.e., the relation between μ_m and $\|m^*_{x|\mu_m}\| = \sqrt{E(m^{*2}_{x|\mu_m})}$ is a rectangular hyperbola. Its verticality or horizontality depends on the sign of $(1 + a_x)c_x - b_x^2$. It can be shown that it is strictly positive so that it is a vertical rectangular hyperbola.[16]

[15] For example, Hou, Xu and Zhang (2020) show that about 67 percent of 447 published anomalies fail to be replicated once the better asset pricing model such as the q-factor model is used as a benchmark. Such results demonstrates the fact that a test on the domestication of a crypto could be very sensitive to the choice of the benchmark asset pricing model.

[16] Σ_x^{-1} is symmetric, thereby having a spectral decomposition

$$\Sigma_x^{-1} = UDU',$$

Equation (10) immediately delivers the global minimum norm such that:

$$\|m_x^*\|^2 = \min_{\mu_m} E\left(m_{x|\mu_m}^{*2}\right) = \frac{(1+a_x)c_x - b_x^2}{(1+a_x)} \text{ where } \mu_m^* = E(m_x^*) = \frac{b_x}{1+a_x}. \tag{12}$$

We can apply the same analysis to the augmented set of assets, x_a. The corresponding square of the minimum-norm of the SDF with the same mean is

$$\inf \|(m_{x_a}|\mu_m)\|^2 = E\left[\left(m_{x_a}^*|\mu_m\right)^2\right] = (1+a_{x_a})\mu_m^2 - 2b_{x_a}\mu_m + c_{x_a}, \tag{13}$$

where

$$a_{x_a} = \mu_{x_a}'\Sigma_{x_a}^{-1}\mu_{x_a}, \quad b_{x_a} = \mu_{x_a}'\Sigma_{x_a}^{-1}\ell_{N+1}, \quad \text{and} \quad c_{x_a} = \ell_{N+1}'\Sigma_{xa}^{-1}\ell_{N+1}.$$

μ_{x_a} and Σ_{x_a} are the $N\times 1$ mean vector and $(N+1)\times(N+1)$ convariance matrix of x_a respectively. Based on the two mean-squared second-norm bounds, we can quantify the upward shift in the frontier induced by the augmentation of the cryptoasset to the stock basis assets.

Proposition 1: *Given μ_m, the difference in the squared second-norm between the minimum-norm SDF of the stock basis assets and its corresponding minimum-norm SDF of the augmented set that share the same mean, μ_m, is*

$$\|m_{x_a|\mu_m}^*\|^2 - \|m_{x|\mu_m}^*\|^2 = \frac{\alpha_{c|\mu_m}^{*2}}{\|e_c\|^2} \tag{14}$$

where

$$\alpha_{c|\mu_m}^* = \left(\mu_c - \beta_c'\mu_x\right)\mu_m - \left(1 - \beta_c'\ell_N\right), \tag{15}$$

and $\|e_c\|^2$ is equal to $\sigma_{e_c}^2$, the variance of residuals, $e_c = x_c - \mu_c - \beta_c'(x-\mu_x)$ given $\beta_c = \Sigma_x^{-1}\Sigma_{xc}$ with $\Sigma_{xc} = E\left[(x-\mu_x)(x_c-\mu_c)\right]$. The two bounds are tangent, i.e.,

$$\|m_{x_a|\mu_m}^*\|^2 - \|m_{x|\mu_m}^*\|^2 = 0 \quad at \quad \mu_m = \frac{1-\beta_c'\ell_N}{\mu_c - \beta_c'\mu_x}.$$

Equation (14) coupled with (15) states that the difference in the squared norm between the two

where the columns of U are orthonormal eignevectors, i.e., $UU' = I_N$ and D is a diagnonal matrix with eigenvalues, $\lambda' = (\lambda_1,\ \lambda_2, \cdots, \lambda_N)$. Again, Σ_x^{-1} is positive definite so that all eigenvalues, $\{\lambda_i\}$, the diagonal terms of D, are strictly positive. Thus, we can rewrite

$$\begin{aligned} a_x &= (U'\mu_x)'D(U'\mu_x) = \lambda'\mathbf{a}_x^2 = \sum_{i=1}^N \left(\sqrt{\lambda_i}\mathbf{a}_{xi}\right)^2 \\ b_x &= (U'\mu_x)'D(U'\ell_N) = \lambda'(\mathbf{a}_x \circ \mathbf{c}_x) = \sum_{i=1}^N \left(\sqrt{\lambda_i}\mathbf{a}_{xi}\right)\left(\sqrt{\lambda_i}\mathbf{c}_{xi}\right) \\ c_x &= (U'\ell_N)'D(U'\ell_N) = \lambda'\mathbf{c}_x^2 = \sum_{i=1}^N \left(\sqrt{\lambda_i}\mathbf{c}_{xi}\right)^2 \end{aligned}$$

where $\mathbf{a}_x = U'\mu_x$ and $\mathbf{c}_x = U'\ell_N$, and $\circ$ is an operator of element-wise product, i.e., a dot product. Therefore, $a_x c_x \geq b_x^2$ from the Cauchy-Schwarz inequality, which results in $1 + a_x c_x - b_x^2 > 0$. Combined with $a_x > 0$, it proves that the right-hand side of Equation (11) is strictly positive. Therefore, when we draw a rectangular hyperbola with μ_m as x variable and $\sqrt{E(m_{x|\mu_m}^{*2})}$ as y variable, it is a 'vertical' rectangular hyperbola.

SDFs is parabolic, which is not surprising because the difference between any two parabolic functions is, in general, parabolic. More importantly, the difference is nonnegative, which is consistent with Lemma 3, which states that the admissible set of SDFs shrinks when the number of assets increases. One interesting finding is that the two bounds coincide at one particular point, $\mu_m = \frac{1-\beta_c'\ell_N}{\mu_c-\beta_c'\mu_x}$. This is a unique feature which emerges when a single asset is added to the basis asset. Figure 4 illustrates these main results of Proposition 1. It shows that adding the cryptoasset to the stock basis assets results in an upward shift in the frontier and the two frontiers meet with tangency at the single point.

Corollary 1: *When an individual cryptoasset is added to the basis assets, the mean-squared second norm bound of the SDF does not move up if the following two conditions are satisfied:*

$$\beta_c'\mu_x = \mu_c \text{ and } \beta_c'\ell_N = 1. \tag{16}$$

The above corollary states that adding the cryptoasset to the basis assets does not result in the upward shift in the frontier if and only if the two conditions are satisfied. Note that β_c is the coefficient of a regression of x_c on x *with an intercept* such that

$$x_c = \beta_{c0} + \beta_c'x + e_c. \tag{17}$$

Consequently, $E(e_c) = 0$ and $\beta_{c0} = \mu_c - \beta_c'\mu_x$. Equation (17) is an outcome of projecting x_c on $(1\ x')'$ as opposed to x such that $P_{(1\ x)}(x_c) = \beta_{c0} + \beta_c'x$. The first condition, $\beta_c'\mu_x = \mu_c$ states that the expected (gross) return on the crypto equals the expected return on the replicating portfolio that is constructed by using β_c as the vector of portfolio weights. This condition is similar to the classic Jensen's alpha, a measurment of pricing errors. However, for it to be so, β_c should be a vector of weights; i.e., its sum should be unity, $\beta_c'\ell_N = 1$, which is the second condition. As such, the second condition seems to be a pre-requisite for the first condition but it is not necessarily so. To see this, suppose the first condition is met, $\beta_c'\mu_x = \mu_c$, which results in $x_c = \beta_c'x + e_c$. x_c is further decomposed into two components:

$$\beta_c'x \in \mathscr{A}(x) \text{ and } e_c \in \mathscr{A}(x)^\perp.$$

That is, the first component belongs to the attainable set of x, i.e., a marketable payoff as defined in Assumption 1. In contrast, the second component, e_c, is orthogonal to x so that it is also orthogonal to $m^*_{x|\mu_m} \in \mathscr{M}(x)$. Therefore, e_c is an idiosyncratic risk component of x_c, which is not priced by $m^*_{x|\mu_m}$. Since $\beta_c'x$ is the only marketable payoff, the price functional dictates $q(\beta_c'x) = \beta_c'\ell_N$. This fundamental price should be one dollar, the market price of x_c. If not, it violates the LOP condition and such a pricing error would be reflected in β_{c0}. In that sense, the first condition seems to be the pre-requisite for the second condition. In sum, the relation between the two conditions is not hierarchical. In addition, a simple interpretation would be such that the second condition measures whether the marketable payoff of the crypto is replicable by the stock basis assets and, as such, it can be regarded as a measure of comovment. In contrast, the first condition states that the replicable component, $\beta_c'x$, is the only priced component. In that sense, it is a measure of price alignment. Overall we can say that $\alpha^*_{c|\mu_m}$ reflects the two domestication traits, comovement and price alignment.

In the previous subsection, we suggest that the minimum norm orthogonal extension, $\varepsilon^*_{c|\mu_m}$ would be the central variable which carries information regarding the two traits that domesticated cryptos should exhibit. Then, a natural question would be how it is related to $\alpha^*_{c|\mu_m}$. That is what we summarize in the following proposition.

Proposition 2: *Given μ_m, $\alpha^*_{c|\mu_m}$ in Proposition 1 is equivalent to the amount of seemingly mispriced value of x_c when $m^*_{x|\mu_m}$ is employed as an SDF, i.e.,*

$$\alpha^*_{c|\mu_m} = E[m^*_{x|\mu_m} \cdot x_c] - 1. \tag{18}$$

Equation (18) can be rewritten as:

$$\alpha^*_{c|\mu_m} = -E[\varepsilon^*_{c|\mu_m} \cdot e_c], \tag{19}$$

where $\varepsilon^*_{c|\mu_m}$ is an orthogonal extension of $m^*_{x|\mu_m}$ which yields the minimum-norm SDF generated by the augmented set of assets, i.e., $m^*_{x_a|\mu_m} = m^*_{x|\mu_m} + \varepsilon^*_{c|\mu_m}$. Specifically, $\varepsilon^*_{c|\mu_m} \perp m^*_{x|\mu_m}$ and $E(\varepsilon^*_{c|\mu_m}) = 0$. In addition,

$$\|\varepsilon^*_{c|\mu_m}\|^2 = \|m^*_{x_a|\mu_m}\|^2 - \|m^*_{x|\mu_m}\|^2 \left(= \frac{\alpha^{*2}_{c|\mu_m}}{\|e_c\|^2} \right) = \inf_{m_{x_a|\mu_m} \in \mathscr{M}(x_a)} \|m_{x_a|\mu_m} - m^*_{x|\mu_m}\|^2. \tag{20}$$

$\varepsilon^*_{c|\mu_m}$ is, in fact, a monotonically transformed variable of the idiosyncratic component of the cryptoasset, e_c:

$$\varepsilon^*_{c|\mu_m} = \frac{-\alpha^*_{c|\mu_m}}{\|e_c\|^2}\, e_c. \tag{21}$$

Equation (21) in Proposition 2 shows that given the mean of the SDF, the minimum-norm orthogonal extension, $\varepsilon^*_{c|\mu_m}$ is perfectly (either positively or negatively depending on the sign of $\alpha^*_{c|\mu_m}$) correlated with the idiosyncratic component, e_c. In addition, rewriting Equation (20) yields

$$\|\varepsilon^*_{c|\mu_m}\| = \frac{|\alpha^*_{c|\mu_m}|}{\|e_c\|}. \tag{22}$$

As such, the second norm of $\varepsilon^*_{c|\mu_m}$ is isomorphic to the information ratio. Note that $\|\varepsilon^*_{c|\mu_m}\|$ refers to the amount of upward shift in the mean-second norm frontier of the SDF given μ_m in Figure fig:geometry1. Equation (22) states that $\|\varepsilon^*_{c|\mu_m}\|$ varies across μ_m entirely due to a change in $\alpha^*_{c|m}$, because $\|e_c\|$ is fixed and is not a function of μ_m. Therefore, the amount of upward shift in the frontier at μ_q in Figure 1 is driven by the amount of seemingly mispriced value designated by $m^*_{x|\mu_q}$. Such a result documented in Proposition 2 coupled with Proposition 1 provides the following important implication:

Proposition 3: *Given a sample, a given cryptoasset's replicability and price alignment can be diagnosed by the following two steps:*

(i) The cryptoasset is replicable by and aligned in price with the stock basis assets if and only if adding the cryptoasset to the stock basis assets does not result in the upward shift of the mean-second norm frontier of the SDF. One can test this by evaluating the two conditions specified in (16) or alternatively by evaluating whether $\|\varepsilon^*_{c|\mu_m}\| = 0$ at 'any' arbitrary μ_m which is different from the estimate of $\overline{\mu}_m$ where $\overline{\mu}_m = \frac{1-\beta_c'\ell_N}{\mu_c - \beta_c'\mu_x}$, at which the two frontiers meet with tangency as shown in Proposition 1.

(ii) *If the mean-second norm frontier moves up, the crypto asset is either not replicable or not aligned in price with the risk stock basis assets. However, it is still replicable by and is aligned in price with the augmented set of basis assets (the stock basis assets+the risk free asset) if the risk-free rate is equal to the* **shadow** *or* **implied** *risk-free rate,* $\overline{\mu}_f = 1/\overline{\mu}_m$.

Proposition 3 summarizes how to evaluate the replicability and the price alignment of a given cryptoasset in the absence of the risk-free asset (step (i)) and in the presence of the risk-free asset (step (ii)). In step (i), the risk-free asset is not included as a basis asset and thus the attainable set spanned by the basis assets is $\mathscr{A}(x)$. As such, the given cryptoasset should be replicated by the stock basis assets only. In contrast, in step (ii), the risk-free asset is included in the set of basis assets, which expands the attainable set to $\mathscr{A}(x_{\mathbb{e}})$ where $x_{\mathbb{e}} = (1/\mu_f x')'$ and μ_f is the risk-free rate. In this case, the cryptoasset is evaluated on whether its payoff is replicable by the stock basis asset and the risk-free asset.

In the absence of the risk-free asset, which is analyzed in step (i), one can choose arbitrarily any value of μ_m as long as it is not $\overline{\mu}_m$. Then examine whether $\|\varepsilon^*_{c|\mu_m}\|$ is zero or equivalently $\alpha^*_{c|\mu_m}$ is zero. If so, the payoff of the given cryptoasset is replicable by the stock basis assets and its price is aligned with them. Geometrically, the mean-second norm frontier does not move up. In the case that $\|\varepsilon^*_{c|\mu_m}\|$ is positive, the cryptoasset is not replicable or/and is not aligned in price with the stock basis assets and geometrically speaking, the frontier does make a shift upward. Simply put, in order to evaluate whether the entire frontier shifts up, all we need to do is to examine whether $\|\varepsilon^*_{c|\mu}\|$ is zero or positive at one single value of μ_m other than $\overline{\mu}_m$.

Even in the case that one finds the cryptoasset is not replicable by or/and is not aligned in price with the stock basis assets in step (i), it is not the end of story. In such a case, there is still a possibility that the cryptoasset is replicable by and is aligned in price with the stock basis assets + the risk-free asset. That is step (ii) and such a possibility can be examined by checking the equality of the shadow interest rate $\overline{\mu}_f$ to the market interest rate μ_f.

Step (i) in Proposition 3 is interesting on its own. In the existing literature on market integration, DeSantis (1995) and Bekaert and Urias (1996) propose a test of integration between two capital markets the intuition behind which is to statistically test the upward shift of the HJ bound when the second market is added to the first market. As shown by Hansen and Jagannathan (1991), there is a duality between the HJ bound without nonnegative restriction and thus the mean-standard deviation frontier of risky assets. DeSantis (1995) and Bekaert and Urias (1996) show that the HJ bound shifts up if and only if the mean-varinance frontier shifts outward to the left and therefore such a market integration test is equivalent to the spanning test of Huberman and Kandel (1987). They evaluate a significant increase in the variance (or the second-norm) of the SDF at two different (and arbitrarily pre-specified) values of μ_m in order to investigate an upward shift of the entire HJ bound.

In contrast, step (i) in Propositon 3 shows that if a single asset rather than multiple assets from the second market is added to a set of assets belonging to the first market, we do not need to evaluate an increase in the variance of the SDF at two different values of of μ_m; an evaluation at a single value of μ_m is sufficient as long as the chosen value of μ_m is different from $\overline{\mu}_m$.

Step (i) in Propositon 3 delivers another critical issue to consider: the difference between $\alpha^*_{c|\mu_m}$ and $\|\varepsilon^*_{c|\mu_m}\|$. Despite their claim to test the shift-up of the HJ bound, DeSantis (1995) and Bekaert and Urias (1996) do not test directly the statistical signficance of an upward shift of the bound. Instead, they test whether the SDFs retrieved from a set of stocks in the first market are able to correctly price a set of stocks in the second market. Setting aside the fact that the integration test utilizes two separate values of μ_m, such a test is equivalent to testing $\alpha_{c|\mu_m} = 0$ instead of $\|\varepsilon^*_{c|\mu_m}\| = 0$. Theoretically speaking, it is true that a test on $\alpha_{c|\mu_m} = 0$ is the same as a test on $\|\varepsilon^*_{c|\mu_m}\| = 0$. However, in terms of test statistics and entailing statistical inference, they are not equivalent. In the presence of multiple assets in the second market (i.e., x_c is a vector or random variables), they test the significance of the estimates of $\alpha_{c|\mu_m}$ using the General Method of Moments so that the weighting matrix of the $\alpha_{c|\mu_m}$ estimates in the test statistic is inversely proportional to the covariance matrix of the estimation errors of $\alpha_{c|\mu_m}$. In contrast, the test statistic of the HJ distance test is the amount of the upward shift of the HJ bound by adopting $E(x_c x_c')^{-1}$ as the weighting matrix.[17] Put simply, the market integration test suggested by DeSantis (1995) and Bekaert and Urias (1996) is analogous to the Hansen-Singleton test, which is a test on $\alpha_{c|\mu_m} = 0$, not the HJ distance test, which is a direct test on $\|\varepsilon^*_{c|\mu_m}\| = 0$.

If we ignore the afore-mentioned difference in test statiscs regarding $\alpha^*_{c|\mu_m} = 0$ and $\|\varepsilon^*_{c|\mu_m} = 0\|$, these two tests are equivalent, theoretically speaking. However, the purpose of our paper is not simply to test whether a particular cryptoasset is domesticated or not. In conjunction with such a 'yes-no' question, we are also interested in a non-polar 'wh' question, i.e., how much it is domesticated or how much it is non-domesticated. Measuring such a degree of domestication of a cryptoasset and comparing it with other cryptoassets in that dimension is another important issue in our paper. $\alpha^*_{c|\mu_m}$ and $\|\varepsilon^*_{c|\mu_m}\|$ may not deliver the same answers to those questions. $\|\varepsilon^*_{c|\mu_m}\| = |\alpha^*_{c|\mu_m}|/\|e_c\|$ is a sort of normalized $\alpha^*_{c|\mu_m}$. Because $\|e_c\|$ differs across the cryptos, $|\alpha^*_{c_1|\mu_m}| > |\alpha^*_{c_1|\mu_m}|$ between two cryptos, c_1 and c_2, does not necessarily mean $\|\varepsilon^*_{c_1|\mu_m}\| > \|\varepsilon^*_{c_2|\mu_m}\|$. This is the central issue to explore below.

(2) Economic Identity of $\alpha^*_{c|\mu_m}$ and $\|\varepsilon^*_{c|\mu_m}\|$

Here we revisit Equation (18) and Equation (19) to better understand the economic meaning of $\alpha^*_{c|\mu_m}$ and $\|\varepsilon^*_{c|\mu_m}\|$. Equation (18) can be interpreted in two extreme ways. One extreme interpretation is that there does not exist such $\varepsilon^*_{c|\mu_m}$ which satisfies Equation (19). This interpretation is equivalent to assuming that $m^*_{x|\mu_m}$ in (18) is a 'sufficient statistic' for the true SDF at least in pricing x_c. Then $\alpha^*_{c|\mu_m}$ can be construed as true mispricing in x_c. The other extreme interpretation is that there exists $\varepsilon^*_{c|\mu_m}$, which can explain the seemingly mispriced portion in the price of x_c, which is $\alpha^*_{c|\mu_m}$. Equation (19) specifies the required property that such $\varepsilon^*_{c|\mu_m}$ should satisfy. That is, Equation (19) is built upon the other extreme assumption that x_c is not mispriced at all! It claims that the seemingly mispriced portion in the price of x_c is nothing but an outcome of the fact that $m^*_{x|\mu_m}$ is not specified sufficiently enough for pricing x_c and therefore one needs to find a missing piece of puzzle, $\varepsilon^*_{c|\mu_m}$ by using a clue, Equation (19).

To see this, let us re-examine Equation (19). Since there could be an infinite random variable which satisifies (19), here we begin with an arbitrary orthogonal extension, $\varepsilon_{c|\mu_m}$. Note that because both $E(\varepsilon_{c|\mu_m})$ and $E(e_c)$ are zero, $E[\varepsilon_{c|\mu_m} \cdot e_c] = \mathrm{Cov}(\varepsilon_{c|\mu_m}, e_c) = \rho(\varepsilon_{c|\mu_m}, e_c)\sigma_{\varepsilon_{c|\mu_m}}\sigma_{e_c}$, where $\rho(\cdot,\cdot)$ refers to the correlation coefficient between two random variables. Then, Equation (19) can be rewritten as:

$$\frac{-\alpha^*_{c|\mu_m}}{\rho(\varepsilon_{c|\mu_m}, e_c)\ \sigma_{e_c}} = \sigma_{\varepsilon_{c|\mu_m}}. \tag{23}$$

[17] See Hansen and Jagannathan (1997) and Jagannathan and Wang (1996).

$\rho(\varepsilon_{c|\mu_m}, e_c)$ is unknown, but Equation (23) dictates $\text{sign}\big[\rho(\varepsilon_{c|\mu_m}, e_c)\big] = -\text{sign}\big[\alpha^*_{c|\mu_m}\big]$. The valid domain of $|\rho(\varepsilon_{c|\mu_m}, e_c)|$ is $(0, 1]$.

Given the domain of $|\rho(\varepsilon_{c|\mu_m}, e_c)|$, the aforementioned two extreme interpretations of $\alpha^*_{c|\mu_m}$ diverge on the assumed value of $\rho(\varepsilon_{c|\mu_m}, e_c)$ in Equation (23).

(a) $\rho(\varepsilon_{c|\mu_m}, e_c) = 0$: In this case, the left-hand-side of Equation (23) explodes to infinity with either sign and Equation (23) does not hold. Or alternatively, there exists a discrepancy between non-zero $\alpha^*_{c|\mu_m}$ (left-hand-side) and zero (right-hand-side) in Equation (19). This implies that $m^*_{x|\mu_m}$ is a sufficient statistic for the true but unobservable SDF. We let m^τ denote the true SDF. If $m^\tau = m^*_{x|\mu_m} + \varepsilon^\tau$ where $\varepsilon^\tau \perp m^*_{x|\mu_m}$ and $\varepsilon^\tau \perp e_c$. $m^*_{x|\mu_m}$, though not a true SDF, is sufficient for pricing x_c since it delivers the same price of x_c that m^τ designates. Then $\alpha^*_{c|\mu_m}$ reflects the true mispricing in x_c.

(b) $\rho(\varepsilon_{c|\mu_m}, e_c) = -\text{sign}[\alpha^*_{c|\mu_m}] \cdot 1$: Given the domain of $\rho(\varepsilon_{c|\mu_m}, e_c)$,

$$\frac{-\alpha^*_{c|\mu_m}}{\rho(\varepsilon_{c|\mu_m}, e_c)\,\sigma_{e_c}} = \sigma_{\varepsilon_{c|\mu_m}} \geq \frac{\big|\alpha^*_{c|\mu_m}\big|}{\sigma_{e_c}} = \sigma_{\varepsilon^*_{c|\mu_m}}, \tag{24}$$

where $\sigma_{\varepsilon^*_{c|\mu_m}}$ is the standard deviation of the the minimum-norm orthogonal extension and thus Equation (24) is precisely the same to the last equation in Equation (22). Equation (24) implies that $\sigma_{\varepsilon^*_{c|\mu_m}}$ is computed based on an assumption that

$$\rho(\varepsilon^*_{c|\mu_m}, e_c) = -\text{sign}[\alpha^*_{c|\mu_m}] \cdot 1. \tag{25}$$

Therefore, $\varepsilon^*_{c|\mu_m}$ is either positively or negatively perfectly correlated with the idioysncratic component of x_c and thus should be a linear transformation of the idioyncratic component such that $\varepsilon^*_{c|\mu_m} = \pi_c \cdot e_c$ with some constant π_c, an amplification factor.[18] Then, $\sigma_{\varepsilon^*_{c|\mu_m}} = |\pi_c|\sigma_{e_c}$ and combining this with (24) yields $|\pi_c| = \frac{1}{\sigma^2_{e_c}}\big|\alpha^*_{c|\mu_m}\big|$. Further Equation (25) implies $\text{sign}[\pi_c] = -\text{sign}[\alpha^*_{c|\mu_m}]$ and accordingly $\varepsilon^*_{c|\mu_m} = -\frac{1}{\sigma^2_{e_c}}\alpha^*_{c|\mu_m} e_c$. This is an alternative derivation of Equation (21) in Proposition 2. In sum, $\varepsilon^*_{c|\mu_m}$ is nothing but a reverse engineerly retrieved orthogonal extension, which can justify the price of x_c under the assumption that x_c is not mispriced at all.

As such, there could be two extreme views on $\alpha^*_{c|\mu}$, depending on the underlying assumption on the value of $\rho(\varepsilon_{c|\mu_m}, e_c)$, either 0 or ± 1. What about $\|\varepsilon^*_{c\mu_m}\|$? Roughly speaking, it is based on an assumption that the tested asset itself, a cryptoasset in our case, comes with its own pricing factor, $\varepsilon^*_{c|\mu_m}$ and that pricing factor is basically its own idiosyncratic payoff, e_c, a residual inexplicable by the stock basis assets. Unfortunately, we cannot tell *a priori* and even *a posteriori* which view is correct simply because we cannot observe the true SDF. Accodingly we also do not have an answer to whether $\alpha^*_{c|\mu_m}$ is a true indicator of mispricing or the remainder of the asset price to be explained by non-stock factors. However, this question is not as critical as it seems in our paper. On the one hand, if $\alpha^*_{c|\mu_m}$ reflects the mispricing, it means that the cryptoasset is not aligned in price with the stocks. On the other hand, if $\alpha^*_{c|\mu_m}$ is driven by non-stock factors, it means that the cryptoasset moves less in tandem with the stocks. As such, a large value of $\alpha^*_{c|\mu_m n}$ means that the cryptoasset is less likely to be domesticated in either way.

[18] A constant term is not necessary in π_c since $E(\varepsilon^*_{c|\mu_m}) = E(e_c) = 0$.

(3) A Unique Feature of the GMN-SDF, m_x^*

We have investigated how the mean-second norm frontier of the SDF behaves when an individual cryptoasset is added to the stock basis assets. In particular, Proposition 3 shows that we can test the upward shift of the frontier by evaluating $\alpha^*_{c|\mu_m} = 0$ or equivalently $\|\varepsilon^*_{c|\mu_m}\| = 0$ at any arbitrary μ_m other than the estimate of $\overline{\mu}_m$. Here we choose $\mu_m^* = E(m_x^*)$, i.e., the mean value of the GMN-SDF, as a natural choice of μ_m. As a matter of fact, there are two good reasons to select μ_m^* as a choice. That is related to how to interpret $\alpha_{c|\mu_m^*}$. From here on, we denote $\alpha^*_{c|\mu_m^*}$ by α_c^* for brevity. The first interpretation is related to $\|\varepsilon^*_{c|\mu_m^*}\|$. As discussed in Proposition 2 and Propositin 3, it measures a distance between the two frontiers under the restriction that the mean value of the SDF is kept the same. In contrast, as will be shown below, the second is associated with $\|\varepsilon_c^*\|$, which refers to the 'minimum' distance between the two frontiers in the absence of such a restriction.

Before elucidating those two interpretations, we first compute the expected value of the expected value of the GMN-SDF. From (12), we know that $\mu_m^* = E(m_x^*) = \frac{b_x}{1+a_x}$. Or alternatively, because $m_x^* = \theta_x' x = \ell_N' E(xx')^{-1}x$, $E(m_x^*) = \mu_x' E(xx')^{-1}\ell_N$.[19]

We begin with the first interpretation of α_c^*. According to Proposition 3, if $\alpha_c^* = 0$, it means that one of two cases happens; the frontier does not move at all upon an addition of the cryptoasset or the frontier moves up but μ_m^* is coincidently identical to $\overline{\mu}_m$. For the latter case, we need to further check whether the shadow risk-free rate, $1/\mu_m^* (= 1/\overline{\mu}_m)$, is equal to μ_f, the prevailing market risk-free rate. In contrast, if $\alpha_c^* \neq 0$, it means that the frontier moves up and we need to check whether $1/\overline{\mu}_m = \mu_f$ only. This two step procedure is somewhat cumbersome particularly when a statistical inference is involved. A good news is that if we can observe the market risk-free rate, we do not need to go through this two step procedure. All we need to do is to evaluate $\alpha^*_{c|\mu_m} = 0$ at $\mu_m = 1/\mu_f$ since it implies that the cryptoasset is replicable by and aligned in price with the stock basis assets or the augmented set of assets (the stock basis assets+the risk-free asset)). If $\alpha^*_{c|\mu_f^{-1}} \neq 0$, the given cryptoasset is not so regardless of whether the set of basis assets includes the risk-free asset or not.

Unfortunately, this straightforward evaluation method in the presence of the risk-free asset is not feasible because the risk-free rate is not constant in the data. In a time-series analysis, we count a sample observation at each point in time, t, as an element of the sample space, Ω, under the assumption that the relevant stochastic processes are stationary. The problem is that the risk-free interest rate such as the interest rate of a money market account, is *locally* risk-free but is time-varying over the observed sample period. As such, the aforementioned simple evaluation method, which requires constant μ_f is not feasible in the data.

As such,we need to come up with an alternative way to implement it in the presence of the time-varying risk-free rate. To do so, we consider the locally risk-free asset, x_f, which is characterized by its first two moments, $E(x_f) = \mu_f$ and $\mathrm{Var}(x_f) = \sigma_f^2$. In addition, we assume $x_f \perp x$ without loss of generality.[20] Then, the extended set of basis assets becomes $x_{\mathbb{e}} = (x_f \ x')'$. Note that this

[19] Using the Sherman-Morrison formula,

$$E(xx')^{-1} = \Sigma_x^{-1} - \frac{\Sigma_x^{-1}\mu_x\mu_x'\Sigma_x^{-1}}{1+\mu_x'\Sigma_x^{-1}\mu_x}.$$

we can prove that $\mu_x' E(xx')^{-1}\ell_N = \frac{b_x}{1+a_x}$.

[20] In that sense, the locally risk-free asset is similar to a zero-beta portfolio in a classic mean-variance analysis, but

extended set is composed of all risky assets, the stock basis assets plus the locally risk-free asset. Let m^*_{e} denote the GMN-SDF retrieved from this extended set of basis assets and let $\mu^*_{m_{\text{e}}}$ denote its expected value.

Proposition 4: *In the presence of the locally risk-free asset with the time-varying interest rate,*

$$\mu^*_{m_{\text{e}}} = \frac{\mu_f + b_x\sigma_f^2}{\mu_f^2 + (1+a_x)\sigma_f^2} \approx \frac{1}{\mu_f} + \upsilon_J, \tag{26}$$

where $\upsilon_J = \frac{b_x\mu_f-(1+a_x)}{\mu_f^3}\sigma_f^2$ is the Jensen's inequality term,. If it is sufficiently small, the seemingly mispriced value of the cryptoasset becomes

$$\alpha^*_{c|\mu^*_{m_{\text{e}}}} \approx (\mu_c - \beta_c'\mu_x)\frac{1}{\mu_f} - (1-\beta_c'\ell_N). \tag{27}$$

Then $\alpha^*_{c|\mu^*_{m_{\text{e}}}} = 0$ (or equivalently $\|\varepsilon^*_{c|\mu^*_{m_{\text{e}}}}\| = 0$) implies that one of the following two conditions are met:

(i) $\alpha^*_{c|\mu^*_{m_{\text{e}}}} = 0$ because $\beta_c'\mu_x = \mu_c$ and $\beta_c'\ell_N = 1$ in Equation (27). The mean-second norm frontier retrieved from the original stock basis assets do not shift up when the cryptoasset is added. The given crypto is replicable by and aligned in price with them.

(ii) $\alpha^*_{c|\mu^*_{m_{\text{e}}}} = 0$ but $\beta_c'\mu_x \neq \mu_c$ and/or $\beta_c'\ell_N \neq 1$ in Equation (27). This implies $\overline{\mu}_m = \mu_f$. The mean-second norm frontier retrieved from the original stock basis assets shifts up when the cryptoasset is added. However, the mean value of the SDF at which the two frontiers meet with tangency is equal to the inversed mean value of the risk-free rate and thus the shadow risk-free rate is the same to the expected value of the risk-free rate. The given crypto is replicable by and aligned in price with the extended set of basis assets (the original stock basis assets + the locally risk-free asset)

If $\alpha^*_{c|\mu^*_{m_{\text{e}}}} \neq 0$ (or equivalently $\|\varepsilon^*_{c|\mu^*_{m_{\text{e}}}}\| \neq 0$), the given cryptoasset is not replicable by or not aligned in price with not only the stock basis assets but also the extended set of basis assets.

If the time-varying risk-free rate is added to the stock basis assets, the expected value of the GMN-SDF is not, in general, exactly the same to the inverse mean of the risk-free rate, because of the Jensen's inequality. The Jensen's inequality term goes away and the expected value of the GMN-SDF is the same to the mean of the risk-free rate if $\mu_f = \frac{(1+a_x)}{b_x}$ or $\sigma_f = 0$. Remember that $\mu^*_m = \frac{b_x}{1+a_x}$ is the mean of the GMN-SDF retrieved from the original stock basis assets. Therefore $\mu_f = \frac{(1+a_x)}{b_x}$, which is equivalent to $\mu^*_m = 1/\mu_f$ implies that the mean value of the GMN-SDF retrieved from the original stock basis assets is the same as that implied by the risk-free rate. In contrast, if σ_f is zero, the risk-free rate does not vary over time, which eliminates the entire Jensen's inequality term. If either of the two conditions are met, the Jensen's inequality term becomes trivial and the expected value of the GMN-SDF retrieved from the extended set of basis assets becomes $1/\mu_f$.[21]

it is a standing-alone individual asset, not a portfolio of risky assets and its time-varyng movement is, in general, extremely limited in a sample, i.e., σ_f is in a vicinity of zero.

[21] In our data, the estimates of $\mu^*_{m_{\text{e}}}$ and $1/\mu_f$ are 0.99911 and 0.99913 respectively before the structural break and 0.99995 and 0.99995 after the structural break. As such, the Jensen's inequality term, υ_J is trivial.

Proposition 4 states that if the Jensen's inequality term is trivial, evaluating $\alpha^*_{c|\mu^*_{m_\oplus}} = 0$ (or equivalently $\|\varepsilon^*_{c|\mu^*_{m_\oplus}}\| = 0$) is 'sufficient' for testing the replicability and price alignment of the given cryptoasset and accordingly its domestication. That is, we do not need to go through the cumbersome procedures described in Proposition 3. The evaluation method proposed in Proposition 4 is a sort of short-cut, which utilizes the property that the mean value of GMN-SDF becomes consistent with the market risk-free rate once we add the locally risk-free asset to the original stock basis assets. As opposed to evaluating separately whether $\overline{\mu}_f = \mu_f$ in Proposition 3, we directly plug μ_f in Equation (27) so that we can indirectly evaluate $\overline{\mu}_f = \mu_f$ by simply testing $\alpha^*_{c|\mu^*_{m_\oplus}} = 0$. As a result, by simply checking $\alpha^*_{c|\mu^*_{m_\oplus}} = 0$, we are enabled to evaluate all at once whether the given crypto is replicable by and aligned in price with the basis assets regardless of whether the set of basis assets includes the risk-free asset or not. Related to it, as mentioned above, in conjunction with testing whether a certain cryptoasset is domesticated, we are also interested in a non-binary question, i.e., how much it is domesticated or non-domesticated. Because $\alpha^*_{c|\mu^*_{m_\oplus}}$ and its corresponding $\|\varepsilon^*_{c|\mu^*_{m_\oplus}}\|$ are a unified test statistic, each of them could be a candidate for domestication measure. In summary, once the locally risk-free asset is incorporated as a component basis asset, the GMN-SDF reveals a unique feature that its accompanying $\alpha^*_{c|\mu^*_{m_\oplus}}$ and $\|\varepsilon^*_{c|\mu^*_{m_\oplus}}\|$ are a composite test statistic and also a potential domestication measure.

Now, let us discuss the second interpretation of α^*_c. Here we assume that there does not exist the risk-free asset. Step (i) in Proposition 3, we fist need to check whether $\mu^*_m = \overline{\mu}_m$. If $\mu^*_m \neq \overline{\mu}_m$, we go on evaluating whether $\alpha^*_c = 0$. What if $\mu^*_m = \overline{\mu}_m$? Then, μ^*_m is not an appropriate choice of μ_m and we have to select other value of μ_m, let's say μ_p, and evaluate whethe $\alpha^*_{c|\mu_p} = 0$. Here the unique feature of μ^*_m kicks in. Even in such a case, we do not need to change μ_m from μ^*_m to μ_p. From Equation (12), we know $\mu^*_m = E(m^*_x) = \frac{b_x}{1+a_x}$ and thus the seemingly mispriced value of x_c is

$$\alpha^*_c = E(m^*_x \cdot x_c) - 1 = (\mu_c - \beta'_c \mu_x)\frac{b_x}{1+a_x} - (1 - \beta'_c \ell_N). \tag{28}$$

What is special about α^*_c? Given that the risk-free asset is assumed not to exist, we need to construct the replicating portfolio by using the stock risky assets only. In such a case, the replicating regression changes from (17) to the following regression without the intercept:

$$x_c = \mathscr{b}'_c x + \omega_c, \tag{29}$$

where $\mathscr{b}_c = E(xx')^{-1}E(x_c x)$ and ω is a residual. Then the following proposition holds.

Proposition 5: *The seemingly mispriced value of x_c against the GMN-SDF is α^*_c in (28) is actually equal to*

$$\alpha^*_c = \mathscr{b}'_c \mathit{1}_N - 1. \tag{30}$$

In addition, an increase in the second-norm of the GMN-SDF upon an addition of the cryptoasset is

$$\inf_{m_{x_a} \in \mathscr{M}(x_a)} \|m_{x_a} - m^*_x\| = \|\varepsilon^*_c\| = \|m^*_{x_a}\| - \|m^*_x\| = \frac{|\alpha^*_c|}{\|\omega_c\|}, \tag{31}$$

where ε^*_c is an orthogonal extension of m^*_x such that

$$\varepsilon^*_c = \frac{1 - \mathscr{b}'_c \ell_N}{\|\omega_c\|^2}\omega = -\frac{\alpha^*_c}{\|\omega_c\|^2} \cdot \omega_c. \tag{32}$$

Equation (30) in Proposition 5 states that the pricing error of x_c against m_x^* as a benchmark SDF is equal to the difference between the price of the replicating portfolio and the price of x_c, where the replicating portfolio is exclusively composed of stock basis assets. Equation (31) shows that $\|\varepsilon_c^*\|$ is the minimum distance between the two mean-second norm frontiers, one retrieved from the stock basis assets and the other retrieved from the stock basis assets + the cryptoasset. In addition, the minimum distance is simply the difference between the minimum norms of the two frontiers and it is a sort of normalized α_c^* where the normalization is made by $\|\omega_c\|$ as opposed to $\|e_c\|$ in Proposition 3. Note that because $E(\omega_c) \neq 0$ in general, ε_c^* is an orthogonal extension of m_x^* without a constraint that the expected value of the extended SDF should be the same to that of the original SDF: i.e., $E(\varepsilon_c^*)$ needs not be zero. In contrast, (21) is an orthogonal extension of m_x^* with such a constraint so that $E(\varepsilon_{c|\mu_m^*}^*) = 0$.

Proposition 5 sheds a new light on how to understand the mean-second norm frontier in the following sense. Firstly, the replicating regression in Equation (29) does not include the intercept term. By doing so, the part of x_c projected by the stock basis assets, $P_x(x_c) = \beta' x$, is enforced to be *a priori* attainable. From the econometric point of view, it may seem to be misspecified since it does not require $E(\omega_c) = 0$. However, the non-trivial value of $E(\omega_c)$ would be reflected in a change in the mean value of the GMN-SDFs from μ_m^* to $\mu_{m_{xa}}^*$. Secondly and more importantly, such an interesting property occurs only when μ_m is equal to μ_m^*, the expected value of the GMN-SDF. That is, (28) is equal to (30) only when $\mu_m = \mu_m^*$. This is a unique feature of the GMN-SDF. We document this feature in Corollary 3.

Corollary 2: *Suppose $\|\varepsilon_c^*\| \neq 0$. It implies that the mean-second norm frontier shifts up. In contrast, if $\|\varepsilon_c^*\| = 0$, it implies either of the following two things happens.*

(i) $\beta_c' \ell_N - 1 = 0$ because $\mu_c - \beta_c' \mu_x = 0$ and $1 - \beta_c' \ell_N = 0$. In this case, adding the cryptoasset to the basis assets does not move the mean-second norm frontier at all.

(ii) $\beta_c' \ell_N - 1 = 0$ because $\frac{b_x}{1+a_x} = \frac{1-\beta_c'\ell_N}{\mu_c - \beta_c'\mu_x}$. In this case, the mean-second norm frontier shifts up but it is tangent at μ_m^*. However, an increase in the frontier at μ_m other than μ_m^* is economically meaningless since it is based on the assumption that the risk-free rate exists and it is $1/\mu_m$. It violates the assumption that the risk-free rate does not exists.

Corollary 3 is important in the following sense. Note that the basis assets are composed solely of risky stocks. Corollary 3 states that in such a case, we do not need to investigate whether the entire mean-second norm bound of the SDF increases with an addition of the cryptoasset. Instead, all we need to do is to investigate whether the second-norm increases at one particular point, $\mu_m = \mu_m^*$. $\|\varepsilon_c^*\| = 0$ is not an equivalent of $\|\varepsilon_{c|\mu_m}^*\| = 0$ for all μ_m because of one exceptional case, i.e., when the two frontiers are tangent at μ_m^*. However, $\|\varepsilon_{c|\mu_m}^*\| > 0$ at μ_m other than μ_m^* is economically meaningless simply because such μ_m is not feasible in the absence of the risk-free asset: Equation (30) is true only at μ_m^*. In sum, if we do not consider the risk-free asset as a part of basis assets, it is sufficient for exploring $\|\varepsilon_c^*\| = 0$.

α_c^* in Proposition 5 is exactly the same to $\phi_c' \ell_N - 1$, which we have seen in Section 4.2 when we discuss the intuition behind economic representation of domestic syndrome. Equation (6) therein is precisely the same to Equation(54) in the proof of Proposition 5. As such, the detailed explanation on how to understand α_c^* can be found in Section 4.2.

Now let us compare Proposition 4 and Proposition 5. To do so, we replace the symbol $x_{\mathrm{e}}' = (x_f \ x')$ by x for a succinct expression from here on. As such, x is redefined as an extended set of basis assets which includes the locally risk-free asset. In Proposition 4, we center upon the role of the locally risk-free asset which enables to simplify the complex evaluation procedure described in Proposition 3. In contrast, the locally risk-free asset in Proposition 5 is construed as nothing but another risky asset added to the original stock basis assets. As such, α_c^* has a dual meaning: it could be interpreted as either Equation (27) in Proposition 4 or Equation (30) in Proposition 5.

α_c^* in Equation (27) is associated with $\|\varepsilon_{c|\mu_m^*}\|$, which is the second norm of the orthogonal extension of the SDF under the restriction that the extended SDF keeps the same mean, i.e., $E(m_{x_a|\mu_m^*}) = E(m_x^*)$, therebye $E(\varepsilon_{c|\mu_m^*}^*) = 0$. In Figure 4, it is illustrated as a vertical distance between the GMN-SDF retrieved from the basis assets, x, and a particular point on the frontier retrieved from the set of assets augmented by the cryptoasset, x_a, where that particular point is on a vertical line extended from the GMN-SDF. According to Proposition 4, $\alpha_c^* = 0$ or equivalently $\|\varepsilon_{c|\mu_m^*}\| = 0$ implies that the given crypto is replicable by and aligned in price with either the original stock basis assets or those added by the locally risk-free asset.

In contrast, α_c^* in Equation (30) is associated with $\|\varepsilon_c^*\|$, which is also the second norm of the orthogonal extension of the SDF, but the extended SDF is not enforced to maintain the same mean, i.e., $E(m_{x_a}^*)$ could be different from $E(m_x^*)$, thereby $E(\varepsilon_c^*)$ not necessarily being zero. In Figure 4, it is depicted as a vertical distance between the two GMN-SDFs. The figure also shows that $\|\varepsilon_c^*\| \leq \|\varepsilon_{c|\mu_m^*}\|$ because $\|\varepsilon_c^*\|$ is the minimum distance between the two frontiers. $\alpha_c^* = 0$ herein or equivalently $\|\varepsilon_c^*\| = 0$ implies that the cryptoasset is replicable by and aligned in price with the combined risky assets, i.e., the stock basis assets plus the locally risk-free asset.

As such, the two versions of interpretation of α_c^* are congruent with each other. Proposition 4 centers upon the unique role of the locally risk-free asset and thus combine the two different cases of $\alpha_c^* = 0$. In contrast, Proposition 5 simply focuses only on the replicability and price alignment of the cryptoasset by the risky assets in the absence of the risk-free asset. Despite this subtle difference, the resulting test statistic is the same, α_c^* but their associated second-norms are different depending on the interpretation of α_c^*.

In summary, the GMN-SDF has a unique feature. If the risk-free asset is considered as a part of the basis assets, we can test whether a given crypto is domesticated simply by evaluating the seemingly misprice value of the crypto or equivalently its corresponding second-norm of the orthogonal extension against the GMN-SDF retrieved from the extended set of basis assets (the stock basis assets + the risk-free asset). If the risk-free asset is not considered as a part of the basis assets, we simply test

(4) Correlation Between $|\alpha_c^*|$ and Idiosyncratic Volatilities

Comovement and price alignment, these two essential traits of domestication may seem to be independent, but it would not likely be the case. In Propositino 4 and Proposition 5, we have seen that α_c^* partially reflects non-replicability of x_c by the stock basis assets or by those augmented by the locally risk-free asset. In addition, as discussed in Section 4.2., another potential *de facto* measure of comovment is the idiosyncratic volatilities, $\|e_c\|$ in Proposition 4 and $\|\omega_c\|$ in Proposition 5. Given the fact that both $|\alpha_c^*|$ and the idiosyncratic volatilities contain the information regarding comovement, they may be positively correlated.

There is another reason for positive correlation between the two. α_c^* also reflects the price misalignment of the crypto with the basis assets. This misalignment melted in α_c^* itself is also highly likely to be correlated with the idiosyncratic volatilities. To see this, we need to get back to the notion of market efficiency. In an efficient market, the market price of an asset should be the unbiased estimate of its fundamental value. The fundamental values are highly correlated among assets in the same market since they are co-affected by common factors governing the market. As a result, their prices should demonstrate strong comovement.

Suppose that in such a market, there exists an asset the price of which deviates from its fundamental value and the scale of deviation is large enough to take advantage of. This means that the price of this asset is derailed from the pricing rule governing the market and its market price is different from a similar asset or portfolio. In such a case, statistical arbitrage would be triggered. Unlike the text-book arbitrage, which relies upon a 'perfect' substitution between the two assets or portfolios and thus is risk-free, statistical arbitrage utilizes a 'sufficient' substitution between them and thus is not risk-free. That is, if a target asset is 'sufficiently' replicable by the other asset or portfolio, their prices should be similar and, if not, statistical arbitrages would seize the opportunity and simultaneously buy the cheap and sell the expensive in the quest for statistical arbitrage gains. Therein, the sufficiency measure of replicability is $\|e_c\|$ or $\|\omega_c\|$, which indicates the amount of remaning risk after replication and that is the very risk that the statistical arbitrager is exposed to. Judgment on whether $\|e_c\|$ or $\|\omega_c\|$ is small enough and thus worthy of implementing the statistical arbitrage is different for different arbitragers, depending on their risk tolerance, in-house capital, funding cost and others. However, the intensity of arbitrage should be, if everything else is the same, disproportional to $\|e_c\|$ or $\|\omega_c\|$. Similar to the text-book arbitrage, this statistical arbitrage transaction *per se* plays a critical role in reducing the mispricing of the target asset and ultimately enhancing market efficiency. Consequently, the remaining pricing error, $\left|\alpha_c^*\right|$, is disproportional to the intensity of arbitrage activity, and, accordingly is porportional to $\|e_c\|$ or $\|\omega_c\|$. This is the reason underlying the positive correlation between the price alignment and comovement.

In a mainstream market, this positive association would be relatively weak. The idiosyncratic risks of assets in this market would be relatively small enough to dampen the statistical arbitrage transactions, and as a result, most of assets herein are correctly priced and their pricing errors are more likely to be random. As such, the pricing errors are crowded around zero with a big mass and only a few assets with exceptionally large idiosyncratic risk are in a blind spot of statistical arbitrage and thus remain to be significantly mispriced. Then, neither of $\left|(\alpha_c^*\right|$ or $\|e_c\|$ (or $\|\omega_c\|$) is cross-sectionally widely dispersed and their correlation would be relatively weak.

In contrast, this positive association should be much more evident in a yet-to-be-mainstream one such as the cryptoasset market. The domesticated cryptoassets, which show comovement in price with the basis assets, i.e., stocks, thereby 'sufficiently' replicable by them, would invite statistical arbitrage and thus their price is more likely to be aligned with the basis assets. In contrast, the cryptoassets that are yet to be domesticated, are not sufficiently replicable by the basis assets and not suitable for statistical arbitrage, and as a result, their market prices substantially deviate from the pricing rule governing the basis market. Therefore, $\|\alpha_c^*\|$ is dispersed as widely as $\|e_c\|$ or $\|\omega_c\|$ and their association would be much more evident in this market.

An important message of Equation (19) is expressed in its alternative form, (24). Remember that (19) is the condition for finding an extension of $(m_x^*|\mu_d)$, i.e., $(\varepsilon|\mu_d)$, under the premise that x_c is not mispriced. This reverse engineering of $(\varepsilon^*|\mu_d)$ is feasible because $x_c \notin \mathscr{A}(x)$, i.e., $e \neq 0$. (24) can be rewritten as:

$$\left|(\alpha_c^*|\mu_d)\right| = \left|E[(m_x^*|\mu_d) \cdot x_c] - 1\right| = \sigma_e \sigma_{(\varepsilon^*|\mu_d)} = \|e\| \|(\varepsilon^*|\mu_d)\|.$$

Therefore, given $\left|(\alpha_c^*|\mu_d)\right|$, the larger $\|e\|$, the smaller $\|(\varepsilon^*|\mu_d)\|$ becomes! Note that $\|e\|$ is exogenously given and $\|(\varepsilon^*|\mu_d)\|$ is endogenously determined. Roughly speaking, $\|\varepsilon^*|\mu_d)\|$ is a measure based on the assumption that the tested asset itself, a crypto-asset in our case, comes with its own pricing factor, $(\varepsilon^*|\mu_d)$, and, more importantly, that pricing factor is its own idiosyncratic payoff, e, a residual inexplicable by the basis assets. Therefore, an asset, which is more difficult to replicate is to match the seemingly mispriced amount with a smaller amplification coefficient, ϕ.[22]

This positive association between α_c^* and $\|e_c\|$ (or $\|\omega_c\|$), if it exists, means that $\|\varepsilon^*_{c|\mu_m^*}\|$ or $\|\varepsilon_c^*$ could be fatally flawed as a measure of domestication. Suppose there are two cryptos with the same amount of α_c^*. Between the two, the crypto-asset with the larger amount of idiosyncratic payoff, $\|e_c\|$ or $\|\omega_c\|$ is assigned with smaller $\|\varepsilon^*_{c|\mu_m^*}\|$ or $\|\varepsilon_c^*$ and thus is determined to be more domesticated. Of course, this is counterintuitive since the larger value of $\|e_c\|$ or $\|\omega_c\|$ implies that cryptoasset is less replicable and thus less comoves with the basis assets.

In sum, regarding the positive relation between $|\alpha_c^*|$ and $\|e_c\|$ (or $\|\omega_c\|$), we can make the following three conjectures.

(1) Weak comovement could be one of important causes for price misalignment by discouraging the statistical arbitrage transactions and as a result, may induce a positive correlation between $\|e_c\|$ (or $\|\omega_c\|$) and $|\alpha_c^*|$.

(2) The distance measures, $\|\varepsilon^*_{c|\mu_m^*}\|$ and $\|\varepsilon_c^*\|$, which measures the amount of upward shift in the mean-second norm frontier of the SDF could be a counterfactual and misleading measure of domestication.

4.4 Candidate Domestication Measures

Based on the above analysis, we consider a set of alternative candidate domestication measures. Combining Equation (22)and Equation (31) yields

$$|\alpha_c^*| = \|e_c\| \cdot \|\varepsilon^*_{c|\mu_m^*}\| = \|\omega_c\| \cdot \|\varepsilon_c^*\|. \tag{33}$$

As discussed above, in Equation (33), $|\alpha_c^*|$, $\|e_c\|$ and $\|\omega_c\|$ are characteristics specific to the cryptoasset. Given these characteristics, $\|\varepsilon^*_{c|\mu_m^*}\|$ and $\|\varepsilon_c^*\|$ are endogenously determined to satisfy the equations in (33). In measuring the domestication syndrome of the given crypto, α_c^* is a natural candidate, which carries information regarding comovement and price alignment altogether. We use α_c^* rather than $\|\alpha_c^*\|$. In the next section, we will nonparametrically construct the distribution of the domestication measures under the null that a given crypto is domesticated. This nonparametrically estimated null distribution may not be symmetric and therefore it could be important to keep the information about the sign of α_c^*.

The next natural candidates are the distance measures, $\|\varepsilon^*_{c|\mu_m^*}\|$ and $\|\varepsilon_c^*\|$. These candidates are attractive in the sense that they measure the amount of upward shift in the mean-second norm frontier. However, as discussed above, they are the variants of $|\alpha_c^*|$ which are normalized by the idiosyncratic volatilities, $\|e_c\|$ and $\|\omega_c\|$. As a result, they may deliver counterintuitive and misleading information regarding domestication if $|\alpha_c|$ and these idiosyncratic volatilities are strongly

[22] This implies that the Hansen-Jagannathan bound, *cetris paribus*, shifts up less when an asset with a larger amount of idiosyncrasy is added to the same basis asset.

correlated. Again rather than use these distance measures as they are, we keep the information regarding the sign of α_c^*. To do so, we slightly modify the distance measures such that

$$\mathrm{sign}[\alpha_c^*] \cdot \|\varepsilon^*_{c|\mu_m^*}\| = \frac{\alpha_c^*}{\|e_c\|} \;\;\&\;\; \mathrm{sign}[\alpha_c^*] \cdot \|\varepsilon_c^*\| = \frac{\alpha_c^*}{\|\omega_c\|}.$$

Recall that α_c^* can be construed as $(\mu_c - \beta_c'\mu_x)(1/\mu_f) - (1 - \beta_c'\ell_N)$ or $b_c'\ell_N - 1$. Given this seemingly mispriced value of the crypto's price, $\|\varepsilon^*_{c|\mu_m^*}\|$ and $\|\varepsilon_c^*\|$ are determined to justify the crypto's market price. As such, it is true that α_c^* determines the two transformed distance measures, not the other way around. Depsite that, let us change the way of thinking such that $\mathrm{sign}[\alpha_c^*] \cdot \|\varepsilon^*_{c|\mu_m^*}\|$ and $\mathrm{sign}[\alpha_c^*] \cdot \|\varepsilon_c^*\|$ are the benchmark measures and α_c^* itself is a transformed variate of these two measures:

$$\alpha_c^* = \vartheta_{ce} \cdot \mathrm{sign}[\alpha_c^*] \cdot \|\varepsilon^*_{c|\mu_m^*}\| = \vartheta_{c\omega} \cdot \mathrm{sign}[\alpha_c^*] \cdot \|\varepsilon_c^*\|,$$

where $\vartheta_{ce} = \|e_c\|$ and $\vartheta_{c\omega} = \|\omega_c\|$. Then, α_c^* can be thought of as a product of an amplification factor (ϑ_{ce} or $\vartheta_{c\omega}$) and the sign-keeping distance measure ($\|\varepsilon^*_{c|\mu_m^*}\|$ and $\|\varepsilon_c^*\|$). These amplification factor is to alleviate the above-mentioned counterintuitive property of the distance measures that the cryptoasset which is more exposed to the idiosyncratic risk is assessed as being more domesticated. We could be more aggressive in dealing with this drawback of the distance measures by increasing $\vartheta_{ce} = \|e_c\|^2$ and $\vartheta_{c\omega} = \|\omega_c\|^2$. Then, the resulting candidate domestication measures are

$$\alpha_c^* \cdot \|e_c\| = \|e_c\|^2 \cdot \mathrm{sign}[\alpha_c^*] \cdot \|\varepsilon^*_{c|\mu_m}\| \;\;\&\;\; \alpha_c^* \cdot \|\omega_c\| = \|\omega_c\|^2 \cdot \mathrm{sign}[\alpha_c^*] \cdot \|\varepsilon_c^*\|.$$

It is true that these candidates seem to be ad hoc. We consider them to see who sensitive the domestication measures are in assessing the domestication of a crypto to the positive relation between $|\alpha_c^*\|$ and the idiosyncratic volatilities.

In summary, we consider the following five candidate domestication measures:

(1) α_c^*

(2) $\frac{\alpha_c^*}{\|e_c\|}$ and $\frac{\alpha_c^*}{\|\omega_c\|}$

(3) $\alpha_c^* \cdot \|e_c\|$ and $\alpha_c^* \cdot \|\omega_c\|$

Which is a more appropriate domestication measure is an empirical question and we examine it in the next section.

5 Measurement of Domestication: Empirical Analysis

5.1 Description of the Data

We use bi-weekly returns of relevant variables over the period from January 2017 to March 2022. We start from January 2017 for three considerations. Firstly, before that date, the number of cryptoassets with reasonable market size and liquidity is too small to empirically assess the domestication of the cryptoasset market. Secondly, the cryptoasset market experiences a number of sudden and structural changes driven by a variety of external shocks and events related to its infra-structure and regulations. With this concern, we avoid going too far back in time in an empirical analysis.

Thirdly, Auer et al. (2022) documents that the institutionalization of the cryptoasset market gained noticeable momentum in the beginning of 2020, which is consistent with the empirical results in Section 3. As such, the post-break time period is relatively short, and we would like to avoid an excess asymmetry in the time span before and after the break.

The descriptive statistics of data is summarized in Table 4. The data covers the time series of biweekly returns over the period from the first week of January 2017 to March 23, 2022. We consider the biweekly returns to obtain a reasonable sample size of time-series data while avoiding a potential 'weekend effect' in stock returns. Consequently, the time-series sample sizes are 82 and 52 before the break and after the break respectively.

We use the Fama and French twenty-five book-to-market and size sorted stock portfolios as a set of stock basis assets, which are downloaded from French's data library. As is well known, these portfolios are common stocks of non-financial corporations listed in NYSE, AMEX, and Nasdaq that are covered by CRSP as well as COMPUSTAT. They are constructed at the end of each June and are the intersections of five portfolios formed on market capitalization and five portfolios formed on the ratio of book to market value. Table 4 shows that after the break, their average returns and corresponding volatilities are slightly higher. We also use the locally risk-free interest rate, which is available on the same library of French as an additional basis asset. The post-break period is mostly covered by the zero-interest rate regime in response to the COVID-19 so that its mean is close to 1, as shown in Table 4.

As will be discussed below, we need returns on individual common stocks to construct a cross-sectional distribution of test statistics under the null that a cryptoasset is domesticated. We use all individual stocks listed in NYSE, AMEX, and Nasdaq that are covered by CRSP. The total number of stocks did not change over the entire period of our empirical analysis thereby remaining constant at 2,649 firms. Their mean and median returns are slightly higher than those of the stock basis assets, the Fama-French twenty five portfolios, whereas their volatilities are far greater than those of the basis assets, which states strong evidence on the idiosyncratic risks.

Following Liu and Tsyvinski (2021) and Liu, Tsyvinski and Wu (2022), we utilize trading data of all cryptoassets available from CoinMarketCap, which is the most-referenced source for cryptoasset price and volume data. For each cryptoasset, CoinMarketCap provides a price, which is the volume-weighted average of all prices reporated in each coin exchange that provides API. Among all the cryptoassets available from it, we consider the cryptoassets that maintain the market capitalizations of more than one million dollars over the relevant sample periods (the pre-break period and the post-break period respectively). As such, our data is composed of the relatively large cryptoassets. We exclude five cryptoassets the price behaviors of which are exceptionally unreliable.[23] The mean and median return of the individual cryptoassets declined slightly along with their volatilities after the break.

One interesting thing to notice in Table 4 is that the cross-sectional maximum volatility of the individual stocks is greater than that of the individual cryptoassets in both of the pre- and post-break periods. Some individual stocks show even a greater amount of fluctuations in their returns than the cryptoassets do.

[23]The five cryptoassets excluded are Mobius, Elamachain, POA network, Pirate Chain and COCOS. Their prices once shot up more than 1,000 percent within two weeks, which were followed by headlong plunges. Excluding them does not change our qualitative conclusion.

5.2 Unbiased Estimators of α_x^*, $\|\omega_c\|$ and $\|e\|$ and Ratios

Before estimating the alternative candidate domestication measures, we first estimate their component variables, α_x^*, $\|e_c\|$ and $\|\omega_c\|$.

Even though consistent estimators are all available in closed-form, we consider small sample estimators under the normality assumption of e_c and correspondingly ω_c with a concern that the size of time-series sample, T, is not large relative to the parameters (N or $N+1$) to be estimated and, as a result, the estimators are highly likely to be subject to a large amount of bias. This problem could be potentially exacerbated by a large amount of residuals, e_c and ω_c in the cryptoasset payoff. As such, we assume

$$e_c \sim N(0, \sigma_{e_c}^2).$$

Under this normality assumption, we compute the unbiased estimators of relevant parameters.

Proposition 6: *α_c^*, the estimator of pricing error of x_c, which is identical regardless of the presence of the risk-free asset, it is distributed as following:*

$$\begin{aligned}\hat{\alpha}_c^* &= \widehat{E(m_x^* \cdot x_c)} - 1 \\ &= \hat{\mathscr{b}}'\ell_N - 1 = \hat{\beta}_c'\ell_N - 1(\hat{\beta}_c'\hat{\mu}_x - \hat{\mu}_c)\frac{\hat{b}_x}{1+\hat{a}_x} \sim \mathcal{N}\left[\alpha_c^*, \frac{\sigma_{e_c}}{T} \cdot \frac{\hat{c}_x + \hat{a}_x\hat{c}_x - \hat{b}_x^2}{1+\hat{a}_x}\right], \qquad (34)\end{aligned}$$

where $\hat{\mathscr{b}}_c$ and $\hat{\beta}_c$ are the ordinary least squares regression coefficients without and with an intercept respectively:

$$\hat{\mathscr{b}}_c = (XX')^{-1}(X'X_c), \quad \hat{\beta}_c = \hat{\Sigma}_x^{-1}\hat{\Sigma}_{xc}.$$

where X and X_c are $T \times N$sample matrix and $T \times 1$ sample vector corresponding to x and x_c and

$$\hat{\Sigma}_x = \frac{X'X}{T} - \frac{X'\ell_T}{T} \cdot \frac{\ell_T'X}{T}, \quad \hat{\Sigma}_{xc} = \frac{X'X_c}{T} - \frac{X'\ell_T}{T} \cdot \frac{\ell_T'X_c}{T}$$

are the sample estimates of a covariance matrix of x and a covariance vector between x and x_c respectively without adjustment for the degrees-of-freedom. Similarly, $\hat{a}_x$, $\hat{b}_x$ and $\hat{c}_x$ are the corresponding sample estimates of a_x, b_x and c_x. N is the number of basis assets including the locally risk-free asset and thus 26 in our sample.

Below, we provide the unbiased estimators of $\|e_c\|$ and its corresponding domestication measures, $\frac{\alpha_c^*}{\|e_c\|}$ and $\alpha_c^*\|e_c\|$ in Proposition 5-1. In contrast, there do not exist the unbiased estimator of ω_c. Instead in Proposition 5-2, we will compute the pseudo-unbiased estimators of $\|\omega_c\|$ and its corresponding domestication measures, $\frac{\alpha_c^*}{\|\omega_c\|}$ and $\alpha_c^*\|\omega_c$. Using an extensive simulation, we demonstrate that the biases of these pseudo-unbiased estimators are trivial. Any letter with a 'hat' denotes an estimator.

Proposition 7-A: *The unbiased estimators of $\|e_c\|$ and its corresponding candidate domestication measures, $\frac{\alpha_c^*}{\|e_c\|}$ and $\alpha_c^*\|e_c\|$ are as following:*

- *The unbiased estimator of $\|e_c\|$ $(= \sigma_{e_c})$:*

$$s_{e_c} = \sqrt{\frac{T-N-1}{2}} \frac{\Gamma\left(\frac{T-N-1}{2}\right)}{\Gamma(\frac{T-N}{2})} \sqrt{\hat{\sigma}^2_{e_c}}$$

where

$$\hat{\sigma}^2_{e_c} = \frac{\sum_{t=1}^{T} \hat{e}^2_{ct}}{T-N-1}$$

is an unbiased estimator of $\sigma^2_{e_c}$ and $\Gamma(\cdot)$ refers to a gamma function.

- *The unbiased estimator of $\frac{\alpha^*_c}{\|e_c\|}$:*

$$\widehat{\left(\frac{\alpha^*_c}{\|e_c\|}\right)} = \frac{\hat{\alpha}^*_c}{\sqrt{\hat{\sigma}^2_{e_c}}} \cdot \sqrt{\frac{2}{T-N-1}} \frac{\Gamma\left(\frac{T-N-1}{2}\right)}{\Gamma\left(\frac{T-N-2}{2}\right)} \overset{or}{=} \frac{\hat{\alpha}^*_c}{s_{e_c}} \cdot \frac{\left[\Gamma\left(\frac{T-N-1}{2}\right)\right]^2}{\Gamma\left(\frac{T-N}{2}\right)\Gamma\left(\frac{T-N-2}{2}\right)}$$

- *The unbiased estimator of $\alpha^*_c\|e_c\|$:*

$$\widehat{\alpha^*_c\|e_c\|} = \hat{\alpha}^*_c s_{e_c}.$$

Similarly, we make the following proposition, which is relevant for $\|\omega_c\|$ and its corresponding candidate domestication measures.

Proposition 7-B: *The pseudo-unbiased estimators of candidate domestication measures and their components in the absence of the risk-free asset are as following:*

- *The pseudo-unbiased estimator of $\|\omega_c\| = \sqrt{E(\omega^2_c)}$:*

$$s_{\omega_c} = \sqrt{\frac{T-N}{2}} \frac{\Gamma\left(\frac{T-N}{2}\right)}{\Gamma(\frac{T-N+1}{2})} \sqrt{\hat{\sigma}^2_{\omega_c}}$$

where

$$\hat{\sigma}^2_{\omega_c} = \frac{1}{1+\hat{a}_x}\hat{\beta}^2_{c0} + \frac{T-1}{T}\hat{\sigma}^2_{e_c}$$

is an unbiased estimator of $E(\omega^2_c)$. $\hat{\beta}_{c0} = \hat{\mu}_c - \hat{\beta}'_c\hat{\mu}_x$, where $\hat{\beta}_c = \hat{\Sigma}^{-1}_x\hat{\Sigma}_{xc}$.

- *The pseudo-unbiased estimator of $\frac{\alpha^*_c}{\|\omega_c\|}$:*

$$\widehat{\left(\frac{\alpha^*_c}{\|\omega_c\|}\right)} = \frac{\hat{\alpha}^*_c}{\sqrt{\hat{\sigma}^2_{\omega_c}}} \cdot \sqrt{\frac{2}{T-N}} \frac{\Gamma\left(\frac{T-N}{2}\right)}{\Gamma\left(\frac{T-N-1}{2}\right)} \overset{or}{=} \frac{\hat{\alpha}^*_c}{s_{\omega_c}} \cdot \frac{\left[\Gamma\left(\frac{T-N}{2}\right)\right]^2}{\Gamma\left(\frac{T-N+1}{2}\right)\Gamma\left(\frac{T-N-1}{2}\right)}$$

- *The pseudo-unbiased estimator of $\alpha^*_c\|\omega_c\|$:*

$$\widehat{\alpha^*_c\|\omega_c\|} = \hat{\alpha}^*_c s_{\omega_c}.$$

Clearly we can see that the unbiased estimators in Proposition 7-A and the corresponding pseudo-unbiased estimators in Proposition 7-B are isomorphic in their functional form except for the degrees of freedom. We examine the simulation analysis to examine the performance of the pseudo-unbiased estimators in Proposition 7-B. The results are reported in Table A.1, which shows that the pseudo-unbiased estimators are 'almost' unbiased. Given this satisfactory performance of the pseudo-unbiased estimators, we use these estimators in Proposition 7-B along with the unbiased estimators in Proposition 7-A in measuring the domestication syndrome for all available crytpo-asset during the pre and post periods of the structural breakpoint, February 24, 2020, which is suggested in the state-space analysis in Section 3.

5.3 Estimation Results of Domestication Statistics

(1) The Relationship Between $\|\alpha_c^*\|$ and Idiosyncratic Risks

Before we compare the estimation results of alternative candidate domestication measures, we first investigate the relationship between the estimates of the absolute seemingly misrprice value ($|\alpha_c^*|$) and the idiosyncratic volatilities ($\|e_c\|$ and $\|\omega_c\|$). As discussed in Section 4, the significan positive correlations would result in a counterintuitive conclusion as regards domestication, when the adjusted distance measures, $\frac{\alpha_c^*}{\|e_c\|}$ and $\frac{\alpha_c^*}{\|\omega_c\|}$, are adopted as domestication measures.

The estimation results before and after the structural break on the individual stocks and the individual cryptoassets are presented in Table 5. Panel (a) summarizes the estimation results with $s_{e_c} = \widehat{\|e_c\|}$ as a regressor whereas Panel (b) reports the estimation results with $s_{w_c} = \widehat{\|\omega_c\|}$ as a regressor. All of the slope coefficients are strongly significant both for the individual stocks and the cryptoassets: i.e., the assets with higher idiosyncratic risk is more likely to show a larger amount of seemingly misprice value. In addition, such a tendency is stronger for the cryptoassets, which is shown in the estimated slope coefficients. For example, Panel (a) shows that as for the cryptoassets, the slope estimates are 0.1791 and 0.1795 before and after the break respectively, whereas they are somewhat smaller, 0.0948 and 0.1096 for the individual stocks. Note also that the estimated slope coefficients are quite similar before and after the break, which demonstrates that such positive associations are relatively stable over time. The adjusted R^2s, the goodness-of-fit statistics, are quite high ranging from 0.2969 to 0.5669. Panel (b) reports the similar results as well.

In summary, we find strong positive relations between $\|\hat{\alpha}_c^*\|$ and the idiosyncratic volatilities, s_{e_c} and s_{w_c}. These positive relations imply that the adjusted distance measures may yield a counterfactual conclusion on the domestication of a cryptoasset.

(2) Two-Step Test of Domestication

Table 6-A and Table 6-B coupled with Table 7-A and Table 7-B summarize the estimation results of alternative domestication measures for domestication. As discussed in Section 4, we consider five candidates for domestication measures. In Table 6-A and 6-B, we report the estimation results of α_c^*, $\alpha_c^*/\|e_c\|$ and $\alpha_c^*\|e_c\|$. For these three statistics, we can test their statistical significance by using the finite-sample distribution derived in Proposition 6 and Proposition 7-A. For example, under the null of $\alpha_c^* = 0$, the estimate of $\alpha_c^*/\|e\|$ is distributed with the student's t; the estimate of α_c^* is normally distributed and $\alpha_c^*\|e_c\|$ is distributed with a product of a normal distribution and a χ distribution. Therefore, these classic parametric tests are feasible, but we do not opt for them herein.

The reason is twofold. Firstly, these tests are not coherent with the objective of the tests, i.e., diagnosing whether a certain cryptoasset is domesticated. Suppose that a certain cryptoasset is rejected in the tests. Such a result is not a conclusive evidence that the crypto is not domesticated. Why? Because even individual stocks, which are domesticated assets by definition may fail to pass these tests. The rejections in these tests are nothing but a statement that the price of a particular asset cannot be explained by the SDF retrieved from the set of basis assets. The rejections do not necessarily imply that the tested asset is different from stocks. Secondly, as discussed in Section 4, the seemingly misprice value, α_c^*s, are not ree from the misspecification errors since they are dependent upon the chosen set of basis assets. We adopt the Fama-French twenty five portfolios augmented by the locally risk-free asset as the basis assets. αs would be different if we choose the different sets of basis assets.

As such, we need an alternative method of testing the domestication of a cryptoasset, which is more suitable for the purpose of the test, and, at the same time, is less exposed to the potential misspecification errors. In this respect, we propose the following novel two-step estimation scheme.

(1) **Construction of a Null Distribution by Using Individual Stocks:** In the first step, we estimate the aforementioned three candidate domestication measures for the entire set of individual stocks covered by the CRSP and construct a cross-sectional distribution of each estimated domestication measure. Since stocks are the benchmark against which an individual cryptoasset's domestimation is measured, this cross-sectional distribution is the nonparametrically estimated distribution defined under the null of domestication.

(2) **Test on Domestication of a Cryptoasset:** In the second step, we investigate the statistical significance of the corresponding estimated measure of a certain individual cryptoasset by testing whether the estimate is inside the confidence interval of the above cross-sectional distribution, given the pre-specified confidence level. If it is inside the interval, the cryptoasset is diagnosed with a 'domesticated' asset and if not, it is diagnosed with a 'non-domesticated' asset.

This test scheme is based on the cross-sectional distribution of the relevant domestication measures of the entire set of stocks. Since stocks are domesticated by definition, this distribution itself is the distribution of test statistics under the null hypothesis that an asset is domesticated. Thus this test scheme is suitable for the purpose of the test. Moreover, any errors inherent in the estimated statistics of the cryptoassets driven by the misspecified basis assets would equally affect the estimated statistics of the individual stocks as well. Therefore, the proposed test scheme would is more robust to the misspecification problem.

In Table 6-A, we report the estimation results before the break. The upper part of the table tabulates the estimated results of the first step, i.e., the individual stocks' cross-sectional distribution of test statics, along with the idiosyncratic risk, $s_{e_c}\left(=\widehat{\|e_c\|}\right)$. The mean value of the idiosyncratic risk, s_{e_c}s, is 7.82%, which is slightly higher than its median, 5.67%, which implies the right-skewness of its cross-sectional distribution. We compute the values of s_{w_c} corresponding to the 95% confidence level (5% p-value), 97.5% confidence level (10% p-value) and 99% confidence level (1% p-value). s_{e_c} itself is not an estimate of the domestication measure. As discussed in Section 4, it is an indirect measure of comovment of an asset with the set of basis assets, Fama-French twenty-five stock portfolios coupled with the locally risk-free asset. The left-end of the confidence interval is not considered since the smaller value of s_{w_c} means extremely strong comovment in price between the asset and the basis assets.

Let us move on to the three candidate domestication measures. The mean value of $\widehat{\frac{\alpha_c^*}{\|e_c\|}}$ is 0.0220, which states that the test stastic is positively baised, albeit slightly. This highlights the strength of our two-step estimation procedure since the ultimate test on the domestication of the cryptoasset in the second step takes into account this bias. This positive bias yields an asymmetry in the confidence interval. For example, the lower end of 95% confidence interval is -0.2343 wheras the corresponding upper end is 0.2544. Subracting the bias from these two ends results in the lower end value of -0.2562 (=-0.2343-0.0220) and the upper end value of 0.2324 (=0.2544-0.0220). The left-skewness of $\widehat{\frac{\alpha_c^*}{\|e\|}}$ is consistent with the above-mentioned right-skewness of the estimate of s_{e_c}. This asymmetry of the cross-sectional distribution rationalizes adopting the raw value of α_c^* as opposed to its absolute value in our test statistics.

Next, let us consider $\hat{\alpha}_c^*$. Since its sign equals that of $\widehat{\frac{\alpha_c^*}{\|e_c\|}}$, $\hat{\alpha}_c^*$ is also slightly positively biased. The shape of its distribution is also asymmetric and overall left-skewed. The distribution of the last measure, $\widehat{\alpha_c^*\|e_c\|}$, is also slightly positively biased and skewed to the left.

The lower part of Table 6-A reports the estimation results of the second step. Firstly, most of the cryptoassets show far higher idiosyncratic risks, s_{e_c}s. Not only its mean is 0.4428, which is 5.7 times greater than the mean of the individual stocks, but also only one single cryptoassets's idiosyncratic risk is inside 95% interval. This result alludes to the fact that the cryptoassets, as a whole, show no sign of comovement in price with the stocks before the structural break. Such excess values of the idiosyncratic risks results in a preposterous inference on domestication when the sign-including distance measure, $\left(\widehat{\frac{\alpha_c^*}{\|e_c\|}}\right)$ is adopted as a measure of domestication. Forty one cryptoassets out of fourty three are inside the 95% confidence interval, i.e., 95.43% of the cryptoassets are evaluated as domesticated assets. When the confidence level is enlarged to 97.5%, all of them are concluded to be 'domesticated.' Such a misleading conclusion delivered by this measure is driven mostly by the excess idiosyncratic risks and also by the positive relation between the estimates of $\hat{\alpha}_c^*$ and those of $s_{e_c}\left(=\widehat{\|e_c\|}\right)$ documented in Table 5.

When $\hat{\alpha}_c^*$ is adopted as a domestication measure, the test results are more in line with what is expected. Only three and nine cryptoassets pass the test of domestication when the confidence level is 95% and 97.5% respectively. Recall that this domestication measure can be thought of as $\|e_c\| \cdot \frac{\alpha_c^*}{\|e_c\|}$, i.e., a variant of the sign-including distance measure which penalizes $\|e_c\|$. Last, accordingly to $\widehat{\alpha_c^*\|e\|}$, which more heavily penalizes the idiosyncratic risk, only one and four cryptoassets are concluded to be domesticated at 95% and 97.5% confidence level. This result reflects the extremely weak comovment in price reflected in the distribution of $\|e_c\|$.

The post-break estimation results are reported in Table 6-B. The distributions of all the statistics of the individual stocks therein are more or less similar to the pre-break counterparts. There are two things to notice though. Firstly, the mean and median values of the s_{e_c} are slightly higher. Seoncdly and more importantly, $\hat{\alpha}_c^*$ are more positively biased. its mean is 0.20%, which is equivalent to 5% per annum, and, as a result, the corresponding confidence intervals are more skewed to the right. This result indicates that the Fama-French stock portfolios coupled with the locally risk-free asset do not well price the individual stocks during this period, and it demonstrates the importance of taking into accout such a bias in evaluating the domestication of the cryptoassets.

Unlike the individual stocks, the average idiosyncratic risk of the cryptoassets has declined during this period. Its mean value is 0.3510, which is far smaller than its pre-break counterpart, 0.4428.

About 18.31% of the cryptoassets show idiosyncratic risk which are below the critical value of 95% confidence level designated by the stocks, which is 0.2164. This percentage increases to 46.02% when the confidence level is enlarged to 97.5%. As such, there is a sizable amount of increase in the percentage of the cryptoassets which comove in price with the stocks. The first domestication measure, $\widehat{\left(\frac{\alpha_c^*}{\|e_c\|}\right)}$, show a counterintuitive result such that the less percentage of the cryptoassets are domesticated after the break. It concludes that 83.86% and 90.60% of the cryptoassets are domesticated when the confidence levels are 95% and 97.5% respectively. These ratios are smaller than their pre-break counterparts, 95.43% and 100%. This counterintuitive results are, of course, not consistent with the empirical results on the structural break on comovement in Section 3.

Next, the second measure, α_c^*, shows that its mean has slightly declined from 0.0709 to 0.0686 after the break. Considering that the $\hat{\alpha}_c^*$ of the individual stocks increased after the break, this downturn of α_c^* of the cryptoassets is in stark contrast. Thus, after the break, the cryptoassets show not only stronger comovment in price with the stock portfolios but also improved price alignment with them! The cryptoassets' upgraded price alignment results in the fact that the ratio of domesticated cryptoassets has soared after the break according to this measure. Specifically, 36.39% annd 50.84% of the cryptoassets are now determined as being domesticated at 95% and 97.5% confidence level respectively, which are far higher than their corresponding figures berfore the break, 6.97% and 20.93%. The last and presumably more conservative measure, $\widehat{\alpha_c^*\|e_c\|}$ also show a similar behavior. 28.43% and 48.91% of the cryptoassets are determined to be domesticated at confidence level of 95% and 97.5% respectively after the break, and these ratios are in sharp contrast with 2.32% and 9.30% before the break. Summing up, we surmise that among the three candidate domestication measures, α_c^* and $\alpha_c^*\|e_c\|$ are more reasonable candidate measures given the fact that their diagnosis on domestication is more aligned with the empirical results of the state-space model explored in Section 3.

Table 7-A and Table 7-B report the estimation results when the idiosyncratic risk is measured by $\|\omega_c\|$ instead of $\|e_c\|$.[24] Not surprisingly, the average and median value of $\|\hat{\omega_c}\|$, the second-norm on the residual on the regression without the intercept are slightly greater than those of $\|\hat{e_c}\|$, the zero mean second-norm on the residual on the regression with the intercept. However, the overall results on determining the domestication of the cryptoassets are qualitatively very similar to those in Table 6-A and 6-B: α_c^* and $\alpha_c^*\|\omega_c\|$ seem to desingate domestication in a more sensible way.

(3) The Determination of Domestication Measures: Correlation Analysis

Herein we investigate which domestication measure is admissible. In Table 6-B and 7-B, we decomposed the whole set of cryptoassets into the set of domesticated cryptoassets and the set of non-domesticated assets designated by each measure over the post-break period.[25] Based on those results, we compute $\rho(m_x^*, m_{dc})$ (the correlation coefficient between the GMN-SDF retrieved from the basis assets and the SDF retrieved from the domesticated cryptoassets) and compare it with $\rho(m_x^*, m_{nc})$ (the correlation coefficient between the GMN-SDF retrieved from the basis assets and the SDF retrieved from the non-domesticated cryptoassets). If the domestication measure is sensible, the SDF retrieved from the domesticated cryptoassets should be highly and positively correlated

[24] Any and every results associated with $\hat{\alpha}_c^*$ in those tables are identical to what are reported in Table 6-A and Table 6-B. We re-report those results for an ease in comparing $\hat{\alpha}_c^*$ with $\widehat{\left(\frac{\alpha_c^*}{\|\omega_c\|}\right)}$ and $\widehat{\alpha_c^*\|\omega_c\|}$.

[25] The number of domesticated assets during the pre-break period is too small to retrieve the SDF. For example, $\hat{\alpha}_c^*$ designate only 3 cryptos as domesticated assets when 95% confidence interval is adopted.

with the GMN-SDF retrieved from the basis assets whereas the correlation between the SDF retrieved from the non-domesticated cryptoassets and GMN-SDF retrieved from the basis assets should be close to zero.

To examine the correlations, there are two issues to resolve. Firstly, we have to decide which SDF of the cryptoassets we have to employ between m^*_{dc} and $m^*_{dc|\mu^*_m}$ for the domesticated cryptoassets (m^*_{nc} and $m^*_{nc|\mu^*_m}$ for non-domesticated cryptoassets). m^*_{dc} (m^*_{nc}) refers to the GMN-SDF retrieved from the domesticated (non-domesticated) cryptoassets. As such, its mean is not necessarily the same to the mean of the GMN-SDF of the basis assets: i.e., $E(m^*_{dc}) \neq E(m^*_x)(= \mu^*_m)$ (and similarly $E(m^*_{nc}) \neq E(m^*_x)(= \mu^*_m)$). In contrast, $m^*_{dc|\mu^*_m}$ ($m^*_{nc|\mu^*_m}$) is the minimum-norm SDF of the domesticated (non-domesticated) cryptoassets where its mean is identical to the mean of the GMN-SDF of the basis assets. Given each domestication measure, we pair each of these two SDFs of the cryptoassets with the GMN-SDF of the basis assets.

There are 5 different domestication measures: α^*_c, $\frac{\alpha^*_c}{\|e_c\|}$, $\alpha^*_c\|e_c\|$, $\frac{\alpha^*_c}{\|\omega_c\|}$ and $\alpha^*_c\|\omega_c\|$. Among them, as shown in Proposition 5, $\frac{\alpha^*_c}{\|\omega_c\|}$ and $\alpha^*_c\|\omega_c\|$ are all based on $\|\omega_c\|$ as the idiosyncratic risk. $\|\omega_c\|$ is a relevant measur of the idiosyncratic risk when we measure the distance between the GMN-SDF of the basis assets and the GMN-SDF of the basis assets augmented by the particular cryptoasset. Accordingly, these two domestication measures, $\frac{\alpha^*_c}{\|\omega_c\|}$ and $\alpha^*_c\|\omega_c\|$, are not built upon a restriction that the two SDFs share the same mean. To be consistent, we pair m^*_{dc} (m^*_{nc}), the GMN-SDF of the domesticated (non-domesticated) cryptoassets designated by these measures with m^*_x, the GMN-SDF of the basis assets. In contrast, $\frac{\alpha^*_c}{\|e_c\|}$ and $\alpha^*_c\|e_c\|$, which are based on $\|e_c\|$, are built upon the restriction that the mean of the SDF of the basis assets augmented by the cryptoasset is the same as that of the GMN-SDF of the original basis assets. Therefore, as for these two candidate measures, we pair $m^*_{dc|\mu^*_m}$ and $m^*_{nc|\mu^*_m}$ with m^*_x. As for α^*_c, we consider both m^*_{dc} and $m^*_{dc|\mu^*_m}$ (and m^*_{nc} and $m^*_{nc|\mu^*_m}$) because α^*_c can be construed as a variant of either of $\frac{\alpha^*_c}{\|\omega_c\|}$ or $\frac{\alpha^*_c}{\|e\|}$ corresponding to $\vartheta = \|\omega_c\|$ and $\|e\|$ respectively as shown in Section 4. All told, we estimate 6 different correlations for each of the domesticated cryptoassets and the non-domesticated measures.

The second issue is how to construct the basis portfolios of the domesticated cryptoassets and the non-domesticated cryptoassets. In estimating m^*_{dc} and $m^*_{dc|\mu^*_m}$ (and m^*_{nc} and $m^*_{nc|\mu^*_m}$), we are not able to use the entire universe of the individual cryptoassets because of difficulty in computing the inverse matrix of their second moments. As shown in Table 6-B and 7-B, the number of the cryptoassets is quite large, ranging from 118 to 395 (63 to 294) for the domesticated cryptoassets (non-domesticated cryptoassets) depending on the domestication measure. As such, we need to construct a set of basis cryptoassets, which is supposed to best represent the whole universe of the cryptoassets. To do this, we use two alternative methods. Firstly, we employ the method of forming basis assets suggested by Ahn, Jennifer and Dittmar (2009). Secondly, we use a bootstrappking analysis. Below we discuss each method and estimation result.

Cluster Analysis

This method uses return correlations to sort individual assets into portfolios based on a cluster analysis. It generates a set of basis assets in which cryptoassets should be highly correlated within groups, but have minimal correlation across. The number of clusters is equivalent to the number of basis assets, and we consider the number of clusters to from 10 to 15.

Table 8 summarizes the estimated correlation coefficients for the domesticated cryptos and the non-domesticated cryptos designated by the alternative domestication measures. Let us begin with $\hat{\alpha}_c^*$. Table 6-B shows that given the 95% confidence interval, for example, $\hat{\alpha}_c^*$ designate 151 cryptos as domesticated assets and the remaining 264 cryptos as non-domesticated assets. We identify 10 clusters out of the 151 domesticated cryptos using the cluster analysis and construct the crypto portfolio corresponding to each cluster, which results in 10 domesticated crypto basis assets. As discussed above, when α_c^* is used as a domestication measure, two SDFs, $m^*_{dc|\mu^*_n}$ and m^*_{dc}, can be considered so that we estimate those two SDFs from the 10 domesticated crypto basis assets. Finally, we compute the correlation coefficients, $\rho(m^*_x, m^*_{dc|\mu^*_n})$ and $\rho(m^*_x, m^*_{dc})$. We go through the same procedure with the 264 non-domesticated cryptos to compute the $\rho(m^*_x, m^*_{nc|\mu^*_n})$ and $\rho(m^*_x, m^*_{nc})$. We do this analysis for a different number of clusters and also for a different confidence level.

Consider the estimation results when the number of clusters is 10. At 95% confidence level, $\rho(m^*_x, m^*_{dc|\mu^*_m})$ is strongly positive, 0.5433. In contrast, $\rho(m^*_x, m^*_{nc|\mu^*_m})$ is quite close to zero, -0.0101. This noticeable difference in correlations between the domesticated cryptos and non-domesticated ones is strong evidence that $\hat{\alpha}_c^*$ is very successful in distinguishing between domesticated cryptos and non-domesticated cryptos, thereby being a strong measure of domestication.

In contrast, when m^*_{dc} and m^*_{nc} are employed as the SDF of cryptoassets, the distinction becomes slightly less evident: 0.4783 vs. 0.0215. In fact, the pair of $m^*_{dc|\mu^*_m}$ and $m^*_{nc|\mu^*_m}$ overall shows a stronger performance than the pair of m^*_{dc} and m^*_{nc}: the correlation of $m^*_{dc|\mu^*_m}$ with m^*_x is stronger than the correlation of m^*_{dc} with m^*_x except when the confidence level is 99%. In addition, all of the correlations between $m^*_{nc|\mu^*_m}$ and m^*_x are trivial and slightly negative whereas the correlations of m^*_{nc} with m^*_x are sometimes quite sizable with positive signs, for example, 0.1883 when the number of clusters is 15 and the confidence level is 99%.

There is another good reason why the pair of $m^*_{dc|\mu^*_m}$ and $m^*_{nc|\mu^*_m}$ is more reasonable. Note that the 95% confidence level is the most conservative criterion in determining the domestication of an individual cryptoasset. Consequently, the SDF of the domesticated cryptoassets based on the 95% confidence level should be most highly correlated with the SDF of stock basis assets, if everything else is the same.[26] The 97.5% confidence level is less conservative in selecting the domesticed cryptos so that the SDF retrieved from the domesticated cryptos based on it is more likely to show the less strong correlation. The 99% confidence level, the least conservative criterion, is more exposed to an error of classifying the non-domesticated assets as domesticated ones and thus its correlation is more likely to be the weakest. In Table 8, the correlations between m^*_{nc} and m^*_x fails to meet this monotonicity condition when the number of basis cryptoassets is small, from 10 to 13. In contrast, $m^*_{dc|\mu^*_m}$ satisfies the monotonicity condition across all clusters. In summary, $\hat{\alpha}_c^*$ demonstrates a very strong performs in distinguishing between the domesticated cryptoassets and the non-domesticated cryptoassets, and the pair of $m^*_{dc|\mu^*_m}$ and $m^*_{nc|\mu^*_m}$ $m^*_c|\mu_m$ seems to be superior to the pair of m^*_{dc} and m^*_{nc} as admissible SDFs of the cryptos.

$\widehat{\left(\frac{\alpha_c^*}{\|e_c\|}\right)}$ and $\widehat{\left(\frac{\alpha_c^*}{\|\omega_c\|}\right)}$, which correspond to the orthogonal extesnsion of the SDF, $\|\varepsilon^*_{c||\mu^*_m}\|$ and $\|\varepsilon^*_c\|$ respectively, show a few weaknesses as a measure of domestication, as expected from the estimation results in Table 6s and Table 7s. Both measures deliver the relatively weak correlations between the SDF of the domesticated cryptos and the SDF of the stock basis assets, ranging from 0.0044 to 0.2545. In particular, $\widehat{\left(\frac{\alpha_c^*}{\|\omega_c\|}\right)}$ yields a counterintuitive behavior of correlations when the number

[26] The more conservative confidence level results in the smaller number of domesticated assets. This may lead to less accuracy in identifying distinct clusters, which may result in weak correlations.

of clusters is small. For example, when the number of clusters is 10 with 99% confidence level, the correlation between the SDF of the domesticated cryptos and the SDF of stock basis assets is 0.0654, which is even lower than 0.1787, the correlation between the SDF of the non-domesticated cryptos and the SDF of the stock basis assets; the SDF of the non-domesticated cryptoassets are more strongly correlated with the SDF of the domesticated cryptos, which is counterintuitive. As for $\widehat{\left(\frac{\alpha_c^*}{\|e_c\|}\right)}$, such a counterintuitive behavior of correlations is not observed. However, the correlations between the SDF of the non-domesticated cryptos and the SDF of the stock basis assets are all negative and, sometimes, with a large magnitude. For example, when the number of clusters is 15 and the confidence level is 95%, the correlation is as low as -0.4043. Note that the non-domesticated cryptoassets are defined as not co-moving whatsoever with the stock basis assets and accordingly not being subject to the pricing rule governing the stock basis assets. Consequently, the SDF of the non-domesticated cryptoassets should be neither positively nor negatively correlated with the SDF of the stock basis assets. This counter-intuitive behavior of correlations between SDFs are consistent with the results in Table 6s and Table 7s, which show that these domestication measures 'over-designate' the cryptoassets as domested assets.

Finally, $\widehat{\alpha_c^*\|e_c\|}$ and $\widehat{\alpha_c^*\|\omega_c\|}$ show the correlation behaviors similar to $\hat{\alpha}_c^*$, but is less capable of distinguishing between the domesticated cryptos and the non-domesticated ones. For example, when the number of clusters is 10 with the 95% confidence level, the correlation of the cryptos designated as domesticated by $\widehat{\alpha_c^*\|\omega_c\|}$ is 0.3841, which is lower than 0.4783, the corresponding correlation when $\hat{\alpha}_c^*$ is adopted as a domestication measure. Similarly, the correlation of $\widehat{\alpha_c^*\|e_c\|}$ is 0.4599, which is also lower than the corresponding figure, 0.5433 when $\hat{\alpha}_c^*$ is adopted as a domestication measure. Overall, the above results indicate that α_c^* is a most successful candidate in distinguishing the domesticated cryptos from the non-domesticated ones.

Bootstrapping Analysis

As an alternative to the cluster analysis, herein we consider bootstrapping in constructing the crypto basis assets. We do this by the following procedures:

(a) We fix the number of crypto basis assets to 25 for each of domesticated cryptos and non-domesticated cryptos. Because each crypto basis asset is a portfolio (i.e., bin in random sampling) composed of individual cryptos, we compute how many of individual cryptos defined to be domesticated (non-domesticated) by a particular domestication measure should be allocated into each portfolio (bin). If the total universe of domesticated (non-domesticated) cryptos are divisible by 25, we end up with equal number of individual cryptos in each crypto basis asset (bin). If not, we compute the quotient and allocate its round-up number of individual cryptos in some basis assets (bins) and its round-down number in the remaining basis assets (bins).

(b) Using the uniform distribution, for a particular crypto basis portfolio (i.e., bin), j, we conduct radom sampling out of the domesticated (non-domesticated) cryptoassets, the number of which is identical to the number pre-specified in step (a). We do this random sampling for each of 25 basis assets, $j = 1, 2, \cdots, 25$. Unlike generic bootstrapping, we conduct the random sampling without replacement to avoid the inclusion of certain individual assets in multiple basis portfolios (bins). Consequently, the resulting 25 crypto basis portfolios are composed of individual cryptoassets which do not overlap across different portfolios.

(c) We costruct the equally weighted portfolio with the individiaul cryptos sampled in (b) for each basis portfolio and compute the relevant SDF out of the 25 randomly constructed basis assets in (c), m^*_{dc} (m^*_{nc}) or $m^*_{dc|\mu^*_m}$ ($m^*_{nc|\mu^*_m}$), depending on the domestication measure. Then we estimate and save the correlation between this crypto-SDF and the SDF of the stock basis assets, i.e., the Fama-French 25 portfolios coupled with the locally risk-free rate.

(d) We iterate the procedures from (a) to (c) for 100,000 times.

As such, we end up with the bootstrapped distribution of correlations between the relevant SDF of the cryptoassets and that of the stocks. Practically speaking, this distribution spans the entire set of correlations that are feasible, when one constructs 25 crypto portfolios. A nice feature of this method is to enable us to test the statistical significance of the correlation since we can compute the p-value of zero correlation relative to the bootstrapped distribution.

Table 9 reports the estimation results. Again there are 6 different correlation distributions each of which corresponds to 5 different domestication measures employed to distinguish between domesticated cryptoassets and non-domesticated ones. Also as in 8., two separate correlation distributions are reported for α^*_c, one for the pair of $m^*_{dc|\mu^*_m}$ and $m^*_{nc|\mu^*_m}$ and the other for the pair of m^*_{dc} and m^*_{nc}. Let us take the case where the domestication measure is $\hat{\alpha}^*_c$ and the pair of the relevant SDFs is $m^*_{dc|\mu^*_m}$ and $m^*_{nc|\mu^*_m}$. This is found to best perform in the cluster analysis reported in Table 8. When the confidence level is 95%, the mean correlation of the bootstrapped distribution is 0.440. Its minimum is as -0.108, whereas its maximum is 0.778. Note that in 6-B, the number of domesticated cryptoassets according to this criterion is 151. Out of 100,000 different ways to construct 25 portfolios out of the 151 cryptoassets classified as domesticated by this criterion, the minimum value of the correlation is -0.108 while its maximum value is as large as 0.778!. The p-value of zero is 0.000, which indicates that the correlation is significantly different from zero. In contrast, the corresponding correlations of the non-domesticated cryptoassets are distributed with mean value of 0.042, which is very close to zero. Its minimum and maximum values are -0.422 and 0.483 respectively, which are quite symmetric in absolute terms. The corresponding p-value of zero correlation is 0.347, which is statistically insignificant. In sum, the SDF of the domesticated cryptoassets is significantly correlated with that of the stock basis assets whereas the SDF of the non-domesticated cryptoassets is not. This result holds when we adopt the less conservative definitions of domestication, i.e., when 97.5% and 95.0% confidence levels are employed. The corresponding p-values of zero for the domesticated cryptos are 0.002 and 0.011 so that the correlation is different from zero at 99% and 99.5% confidence level respectively. In addition, the mononicity of the correlation with the confidence level holds here as well. With 97.5% confidence level, the mean value of correlation, 0.348, which is lower than 0.440 when the confidence level is 95.0%, but it is higher than 0.268 when the confidence level is 99.0%. All of the SDFs of the non-domesticated cryptoassets are statistically insignificantly correlated with the SDF of the stock basis assets: the p-values are 0.359 and 0.555 for 97.5% confidence level and for 99% confidence level respectively. All in all, we can conclude that $\hat{\alpha}^*_c$ is an admissible domestication measure that can distinguish between the domesticated cryptoassets and the non-domesticated cryptoassets. This result is consistent with the estimation results delivered by the cluster analysis documented in Table 8.

The implication of the other domestication measures are quite similar to what we observed in Table 8. The estimates of $\alpha^*_c\|e_c\|$ and $\alpha^*_c\|\omega_c\|$ show a solid performance in differentiating the domesticated cryptos form thenon-domesticated cryptos but not as good as $\hat{\alpha}^*_c$. In addition, the performance of the estimates of $\frac{\alpha^*_c}{\|e_c\|}$ and $\frac{\alpha^*_c}{\|\omega_c\|}$ are not satisfactory at all. All the p-values of zero correlation is greater than 5% for the domesticated cryptos. The poor performance of these domestication measures are consistent with the results in Table 8.

(4) Nonparametric SDFs and Stock/Crypto Factors

Here we investigate the relations between the SDFs of the cryptos and the stock market factors along with the crypto-market factors that are proposed in the existing asset pricing literature. As for the stock market factors, we consider two alternative sets of factors: the Fama-French five factors proposed in Fama and French (2015) and the q^5 factors of Hou et al. (2021). The Fama-French five factors are composed of market factor (MRMF), size factor (SML), book-to-market factor (HML), profitability factor (RMW) and investment factor (CMA). In contrast, the q^5 factors are market factor (R_M), size factor (R_ME), investment factor (R_IA), profitability factor (R_ROE), and expected growth factor (R_EG). The name of some factors are common between the two models, but the exact procedures of constructing those factors are different.

As for the crypto-market factors, Liu, Tsyvinski and Wu (2022) document that three factors, cryptoasset market factor (CMKT), size factor (CSMB) and momentum factor (CMOM) are able to capture the cross-sectional dispersion in expected returns on the cryptoassets. Following their study, we replicate those three crypto-market common factors with our sample data.

In the previous subsection, we find that α_c^* is the best performing measure of domestication. In addition, the pair of $m^*_{dc|\mu^*_m}$ and $m^*_{nc|\mu^*_m}$ are the better SDFs than the pair of m^*_{dc} and m^*_{nc}. Thus, we report the regression results on this pair of SDFs.[27]

We consider the following regressions:

$$\begin{aligned}(m^*_{dc|\mu^*_m})_t &= \vartheta_d + \vartheta'_{s,d} f_{st} + \vartheta'_{c,d} f_{ct} + \epsilon_{dt} \\ (m^*_{nc|\mu^*_m})_t &= \vartheta_n + \vartheta'_{s,n} f_{st} + \vartheta'_{c,n} f_{ct} + \nu_{nt}.\end{aligned}$$

where Recall that $(m^*_{dc|\mu^*_m})_t$ is the SDF of the cryptoassets designated as domesticated by $\hat{\alpha}_c^*$ and $(m^*_{nc|\mu^*_m})_t$ is the SDF of the cryptoassets designated as non-domesticated by $\hat{\alpha}_c^*$. Both SDFs are retrieved from the 10 crypto basis assets that are constructed by the cluster analysis in Table 8. The mean values of both SDFs are equal to m_x^*, the GMN-SDF of the stock basis assets, i.e., the Fama-French 25 portfolios augmented by the locally risk-free asset. The regressors, i.e., the stock market and crypto-market factors are

$$\begin{aligned} f_{st} &= \begin{cases} \text{MRMF, SML, HML, RMW, CMA} & \text{for the Fama-French model} \\ \text{R\_MKT, R\_ME, R\_IA, R\_ROE, R\_EG} & \text{for the } q^5 \text{factor model} \end{cases} \\ f_{ct} &= \text{CMKT, CSMB, CMOM.} \end{aligned}$$

Table 10s and Table 11s document the estimation results. In Table 10-A, we present the regression results for domesticated cryptoassets for different confidence levels adopted in defining domestication, where the Fama-French factors are employed as the stock market factors. When the 95% confidence level is adopted, none of the coefficients associated with the three crypto-specific factors are statistically significant either when they are stand-alone factors or when they are comingled with the stock-specific factors. In contrast, two stock factors, the stock market portfolio (MRMF) and the profitability factor (CMA) are statistically significant with the p-values of less than 1% and

[27] The regression results on m^*_{dc} and m^*_{nc}, the GMN-SDF of the cryptos are qualtitatively similar and they are available upon request.

2.5% respectively. The adjusted R^2 is negative in Regression (1), which uses the crypto-market factors as the regressors, whereas Regression (2), which use the stock market factors as the regressors shows the higher adjusted R^2 than Regression (3), which are based on both sets of factors. This result is remarkable. Unlike stocks and other securities, the cryptoassets are devoid of future cash flow and thus, fundamentally speaking, their prices do not have any reason to be affected by the stock-specific factors. Consequently, the only plausible explanation of these empirical results is the non-fundamentaly driven correlations. In particular, our results deliver strong evidence supporting the category view proposed in Barberis and Shleifer (2003), which aruges that the cryptoassets comove with the stocks in price just because a large community of institutional investors began to incorporate some of cryptoassets at par with stocks in their portfolios.

As the domestication criterion becomes more lenient, the aforementioned implication becomes weaker because less domesticated cryptoassets are classified as domesticated. For example, when 97.5% confidence level is adopted, the crypt size factor (CSMB) and the crypto momentum factor earn the marginally significant explanatory power in Regression (6), which uses both of the stock market factors and the crypto-market factors as independent variables. Consequently, the adjusted R^2 of Regression (6) is greater than that of Regression (5), which relies upon the stock market factors as regressors. However, up to this level of confidence level, the stock market factors dominate the the crypto market factors by a large margin in explaining the domesticated cryptos, which can be seen in their adjusted R^2. Such a tendency becomes more noticeable as the most loose criterion of domestication, i.e., 99% confidence level, is adopted; the statistical significance of the crypto size factor (CSMB) and the crypto momentum factor (CMOM) becomes stronger. At this level of confidence level, the crypto-market factors become more powerful than the stock market factors in explaining the domesticated cryptoassets, albeit by a small margin.

Table 10-B documents the regression results for non-domesticated cryptoassets. When the confidence level is 95%, the crypto size factor is the only significant factor in Regression (3). In terms of adjusted R^2, Regression (1) dominates Regression (3), not to mention Regression (2), the R^2 of which is negative. At 97.5% confidence level, all the three crypto-market factors become significant. This result makes sense. As the domestication criterion becomes more lenient, the set of non-domesticated cryptos shrinks and the less strongly non-domesticated cryptos are likely to be excluded. As a result, the degree of non-domestciation becomes stronger as the domestication criterion becomes weaker. Such intuition appears to break down when we move to 99% confidence level, which show that besides the crypto-market factors, the three stock market factors, SMB, HML, and RMW, regain statistical significancy. However, the most likely cause of such counterintuitive results is multicoliniarity between the set of crypto-market factors and the stock market factors. In Regression (8), RMW is the only significant stock market factor and its adjusted R^2 is as small as 5.97%. Adding the stock market factors to the crypto market factors increases adjusted R^2 only by 0.77%, from 49.04% to 49.81%. When comparing the adjusted R^2 of the regression equation based on the stand-alone crypto market factors, Regression (1), Regression (2) and Regression (3), the adjusted R^2 monotonically increases from 12.14% to 38.99% and 49.04% as the less lenient domestication criterion (and thus the more strong non-domestication criterion) is adopted.

Table 11-A and Table 11-B report the estimation results equivalent to Table 10-A and Table 10-B. Here the regression employes the q^5 factors as representive stock market factors in lieu of the Fama-French factors. Overall, the results are quite similar to those with the Fama-French factors. Specifically, in Table 11-A, two stock market factors, the market factor (R_MKT) and the investment factor (R_IA) are statistically significant in explaining the domesticated cryptos. This result is remarkably consistent with that in Table 10-A, which documents the statistical significance

of the market factor (MRMF) and the investment factor (CMA). The exact way of constructing those factors are slightly different but R_MKT and R_IA are qualitatively isomorphic to MRMF and CMA respectively. In addition, at 95% confidence level, no single crypto-market factor survived in statistical significance and as the domestication criterion weakens, CSMB begins to earn statistical significance in Regression (4) and (5), which is also consistent with Table 10-A. In contrast, Table 11-B shows that the stock market factors fail to explain the non-domesticated cryptos except that R_ME is marginally significant in stand-alone regressions, Regression (3) and Regression (4). In addition, all the three crypto-market factors show strong performance in explaining the SDF of the non-domesticated cryptos in the presence of the stock market factors, particularly when the confidence level is 97.5% and 99%.

All in all, the estimation results in Table 10s and Table 11s reaffirm the fact that α_c^* is a reliable measure of domestication that distinguishes between the domesticated cryptos and the non-domesticated cryptos.

(5) Robustness Tests

The outcomes presented in our study are rooted in the Fama and French size-book to market sorted 25 portfolios as the basis asset for the equity market. Nonetheless, as pointed out by Lewellen, Nagel and Shanken (2010), the Fama and French double-sorted portfolios exhibit a strong factor structure, leading to a contraction in pricing errors. In response, we undertake, within this subsection, two supplementary tests to assuage this concern. Initially, we explore alternative Fama and French double-sorted portfolios, namely size and operating profit, size and investment portfolio (which form the Fama and French five-factor model), and size and momentum double-sorted portfolios. Each of these serves as a candidate for the basis asset, and we rerun the entire procedure for each choice. In the Internet Appendix, we provide details on the extent of domestication and the behavior of corresponding stochastic discount factors. While the degree of domestication varies with the choice of the basis asset, our key observation remains significant—the proportion of domesticated crypto-assets substantially increases around the structural breakpoint. Notably, determining domestication based on $\hat{\alpha}_c^*$ is found to be a judicious method regardless of the basis asset chosen.

In parallel, we construct the pricing kernel using the risk-premium principal component analysis (RPPCA) methodology as advocated by Lettau and Pelger (2020). Specifically, we gather 35 anomalies from Chen and Zimmermann (2022) and estimate five generalized principal components, accounting for risk premiums.[28] Subsequently, we compute the implied stochastic discount factor (SDF) as

$$M_t = 1 - \hat{\varphi}'(\hat{F}_t - E[\hat{F}_t]), \text{where } \hat{\varphi} = \Sigma_F^{-1}\mu_F$$

Following this, $\hat{\alpha}_c^*$ is obtained for individual stocks and coin assets, and the entire two-step domestication test is repeated. The design of the risk-premium principal components ensures correlations close to zero, thereby ensuring that the ensuing results are devoid of a pronounced factor structure. Table 12-A summarizes the estimation outcomes. The upper panel displays pre-break results, comparable to those in Table 6-A, indicating a consistent level of domestication using the SDF implied by RPPCA. In the lower panel, post-break results reveal a substantial increase in overall domestication after structural breakpoints. Based on a 95% confidence level, only 11.63% of crypto

[28] The list of 35 anomalies is available in the Internet Appendix.

assets are identified as domesticated before the break, rising to 50.84% afterward—surpassing the original figure of 36.39% from Table 6-B.

We demonstrate the robustness of the overall level of domestication to changes in basis assets. Additionally, we explore the behavior of domesticated and non-domesticated pricing kernels. Table 12-B illustrates the correlation between the implied pricing kernel by RPPCA and the stochastic discount factor based on the cluster analysis presented in Table 8. For domesticated crypto-assets, the overall correlation is slightly lower than that in Table 8 but remains positively significant. Non-domesticated coin assets exhibit correlation close to zero. We employ a bootstrapping procedure for both domesticated and non-domesticated coin assets, with Table 12-C showcasing correlation results derived from bootstrapping. Similar to the findings in Table 12-B, correlations for domesticated coins decrease slightly compared to the corresponding Table 9. Nevertheless, p-values at all confidence levels remain lower than typical thresholds. In summary, we assert the resilience of our results to variations in basis assets, and findings based on RPPCA align seamlessly with our primary results, underscoring the stability of our empirical outcomes.

6 Conclusion

There is nothing geometric about the fact that the crypto-market became institutionalized after many twists and turns in its relatively short history. The institutionalization signifies that this new set of investable assets are incorporated into the portfolios of major institutional investors in conjunction with mainstream assets such as stocks. As a consequence, it should give rise to an inevitable impact on the prices of the cryptos. Then a natural question that would follow is whether they comove in price with the mainstream assets and, if so, the pricing rule governing the institutionalized cryptoassets is congruent with that governing the mainstream assets. These are the questions addressed in this paper.

To answer the above questions, we need to construct a reliable measure which captures the degree of domestication syndrome. Among a number of candidates, an extensive set of tests for horse racing gives us the same message. α_c^* is the legitimate measure of domestication and the best performing measure. According to this domestication measure, the proportion of the domesticated cryptos has skyrocketed from 7% to about 36% after the break, February 2020, when we apply a conserative domestication criterion, 95% confidence level. Interestingly this result is consistent with the dramatic increase in the institutionalization of cryptoassets documented in the survey literature.

However, it is pre-mature to conclude that the domestication process of cryptoassets will keep moving on. Domestication process is topsy-turvy in every dimesion. During the 19th century, colonists tried to domesticate zebras during the trips to Africa, only to fail. Once they were thought to be reasonably domesticated, their dormant wileness appeared to surface. In addition, even formerly domesticated organisms escape controlled cultivation and become easily feralized. As such, it is fair to say that the domestication process of the cryptoassets is in a precarious position at this stage.

Our analysis has implication for practical applications . Some investors may prefer domesticated cryptos as investable instruments to enhance the diversification of their portfolios while keeping away from crypto-specific risks they are not familiar with. Contrariwise, other investors may favor yet-to-be domesticated cryptos in search of extra returns that they cannot expect in the mainstream assets. α_c^* and its corresponding diagnostic method developed in this paper can be used for both of these diametrically opposite types of investors.

A Proofs

Proof of Lemma 1: See Hansen and Jagannathan (1991) for the formal proof. Here for completion, we deliver a less formal proof. The Riesz Representation Theorem states that $\mathscr{M}(x)$ is non-empty and that there is a unique square-integrable random variable, m_x^*, in $\mathscr{M}(x) \cap \mathscr{A}(x)$. Since $m_x^* \in \mathscr{A}(x)$, the projection theorem yields $m_x^* = \theta_x' x$. Then $E(m_x \cdot x) = E(m_x^* \cdot x) + E(\varepsilon_x \cdot x) = \ell_N$ since $E(\varepsilon_x \cdot x) = 0$ for all $\varepsilon_x \in \mathscr{A}(x)^{\perp}$. $\mathscr{M}(x) - \{m_x^*\} \neq \emptyset$ because $\mathscr{A}(x)^{\perp} \neq \emptyset$, which comes from the market incompleteness assumed in Assumption 2. $\|m_x^*\| = \min \|m_x\|$ simply due to $\|m_x\|^2 = \|m_x^* + \varepsilon_x\|^2 = \|m_x^*\|^2 + \|\varepsilon_x\|^2 < \|m_x^*\|^2 \; \forall m_x \in \mathscr{M}(x) - \{m_x^*\}$, given the orthogonality of ε_x. *q.e.d.*

Proof of Lemma 2: $A(x_a) = \{y \in \mathscr{L}^2$ such that $y = \gamma' x_a, \; \gamma \in \mathbb{R}^N\}$. Since any payoff $y = \theta' x \in \mathscr{A}(x)$ can be rewritten as $y = \gamma' x_a$ where $\gamma = (\theta' \quad 0)'$, thereby $y \in \mathscr{A}(x_a)$. This proves $\mathscr{A}(x) \subset \mathscr{A}(x_a)$. In contrast, in $\mathscr{A}(x_a)$, there exists $y = \gamma' x_a$ where the last element of γ_a, i.e., the dollar investment on x_c, is non-trivial. Such $y \notin \mathscr{A}(x)$, and therefore $\mathscr{A}(x) \not\supset \mathscr{A}(x_a)$. *q.e.d.*

Proof of Lemma 3: $\mathscr{M}(x_a) = \{m_{x_a} | E(m_{x_a} \cdot x_a) = \ell_{N+1}\} = \left\{m_{x_a}^* + \varepsilon_{x_a} \middle| \varepsilon_{x_a} \in \mathscr{A}(x_a)^{\perp}\right\}$, where $m_{x_a}^*$ is the minimum-norm stochastic discount factor which can correcly price x_a. Following Lemma 1, $m_{x_a}^* = \theta_{x_a}' x_a$ where $\theta_{x_a} = E(x_a x_a')^{-1} \ell_{N+1}$. Since it correctly prices x as well, there exists a unique $\varepsilon_{x_a}^*$ such that $m_{x_a}^* = m_x^* + \varepsilon_{x_a}^*$ and $\varepsilon_{x_a}^* \in \mathscr{A}(x)^{\perp}$ from Lemma 1. Lemma 2 shows that $\mathscr{A}(x) \subset \mathscr{A}(x_a)$ and $\mathscr{A}(x) \not\supset \mathscr{A}(x_a)$. Therefore $\mathscr{A}(x)^{\perp} \supset \mathscr{A}(x_a)^{\perp}$ and $\mathscr{A}(x)^{\perp} \not\subset \mathscr{A}(x_a)^{\perp}$. $\varepsilon_x \in \mathscr{A}(x)^{\perp}$ and $\varepsilon_{x_a} \in \mathscr{A}(x_a)^{\perp}$ for any ε_x in $\mathscr{M}(x)$ and any ε_{x_a} in $\mathscr{M}(x_a)$ respectively, which yields the desired results. *q.e.d.*

Proof of Lemma 4: The proof here is also from Hansen and Jagannathan (1991). Given μ_m, we augment the set of basis assets by adding $x_f = 1/\mu_m$ as an additional basis asset. Denote the payoff vector of this augmented set of basis assets by $x_{\mathbb{e}}' = (x_f \quad x')$. Correspondingly, we re-define the attainable set as $\mathscr{A}(x_{\mathbb{e}})$ and the set of admissible SDFs as $\mathscr{M}(x_{\mathbb{e}})$. The existence of $m_{x|\mu_m}^*$ is a direct result of the Riesz Represen Theorem as shown in the proof of Lemma 1. Since $m_{x|\mu_m}^*$ is marketable and its mean is μ_m, it should be in the form of $\mu_m + \theta_x'(x - \mu_x)$. To get the solution to θ_x, we insert this trial solution to $m_{x|\mu_m}^*$ into the fundamental valuation equation of the stock basis assets:

$$\begin{aligned} \ell_N &= E\left[(\mu_m + \theta_x'(x - \mu_x)) \cdot x\right] \\ &= \mu_m \mu_x + \Sigma_x \theta_x, \end{aligned}$$

which yields $\theta_x = \Sigma_x^{-1}(\ell_N - \mu_m \mu_x)$ and, as a result, $m_{x|\mu_m}^* = \mu_m + (\ell_N - \mu_m \mu_x)' \Sigma_x^{-1}(x - \mu_x)$. Then its second-norm is

$$\|m_{x|\mu_x}^*\|^2 = E(m_{x|\mu_x}^*)^2 + \sigma^2(m_{x|\mu_x}^*) = \mu_m^2 + (\ell_N - \mu_m \mu_x)' \Sigma_x^{-1} (\ell_N - \mu_m \mu_x) \tag{35}$$

Simplifying Equation (35) yields Equation (10). *q.e.d.*

Proof of Proposition 1: From (10) and Equation (13), given an arbitrary value of μ_m, we can compute the difference in the minimum second-norm between the two SDF:

$$E\left(m_{x_a|\mu_m}^{*2}\right) - E\left(m_{x|\mu_m}^{*2}\right) = \nu_a \mu_m^2 - 2\nu_b \mu_m + \nu_c, \tag{36}$$

where $\nu_a = a_{x_a} - a_x$, $\nu_b = b_{x_a} - b_x$ and $\nu_c = c_{x_a} - c_x$. Let $\sigma_c^2 = E[(x_c - \mu_c)^2]$. Then the inverse of covariance matrix of the augmented payoff is

$$
\begin{aligned}
\Sigma_{x_a}^{-1} &= \begin{bmatrix} \Sigma_x & \Sigma_{xc} \\ \Sigma_{xc}' & \sigma_c^2 \end{bmatrix}^{-1} \\
&= \begin{bmatrix} \Sigma_x^{-1} + \Sigma_x^{-1}\Sigma_{xc}\left(\sigma_c^2 - \Sigma_{xc}'\Sigma_x^{-1}\Sigma_{xc}\right)^{-1}\Sigma_{xc}'\Sigma_x^{-1} & -\Sigma_x^{-1}\Sigma_{xc}\left(\sigma_c^2 - \Sigma_{xc}'\Sigma_x^{-1}\Sigma_{xc}\right)^{-1} \\ -\left(\sigma_c^2 - \Sigma_{xc}'\Sigma_x^{-1}\Sigma_{xc}\right)^{-1}\Sigma_{xc}'\Sigma_x^{-1} & \left(\sigma_c^2 - \Sigma_{xc}'\Sigma_x^{-1}\Sigma_{xc}\right)^{-1} \end{bmatrix} \\
&= \begin{bmatrix} \Sigma_x^{-1} + \frac{\beta_c\beta_c'}{\sigma_{e_c}^2} & -\frac{\beta_c}{\sigma_{e_c}^2} \\ -\frac{\beta_c'}{\sigma_{e_c}^2} & \frac{1}{\sigma_{e_c}^2} \end{bmatrix}
\end{aligned}
$$

Note that we use $\beta_c = \Sigma_x^{-1}\Sigma_{xc}$ and $\sigma_{e_c}^2 = \sigma_c^2 - \beta_c'\Sigma_x\beta_c = \sigma_c^2 - \Sigma_{xc}'\Sigma_x^{-1}\Sigma_{xc}$. Therefore, we can rewrite a_{x_a} as following:

$$
\begin{aligned}
a_{x_a} &= \mu_{x_a}'\Sigma_{x_a}^{-1}\mu_{x_a} \\
&= (\mu_x' \; \mu_c)\Sigma_{x_a}^{-1}\begin{pmatrix} \mu_x \\ \mu_c \end{pmatrix} \\
&= \mu_x'\left(\Sigma_x^{-1} + \frac{\beta_c\beta_c'}{\sigma_{e_c}^2}\right)\mu_x - 2\frac{\mu_x'\beta_c}{\sigma_{e_c}^2}\mu_c + \frac{\mu_c^2}{\sigma_{e_c}^2} \\
&= \mu_x'\Sigma_x^{-1}\mu_x + \frac{1}{\sigma_{e_c}^2}\left[\left(\beta_c'\mu_x\right)^2 - 2\left(\beta_c'\mu_x\right)\mu_c + \mu_c^2\right] \\
&= a_x + \frac{1}{\sigma_{e_c}^2}\left(\mu_c - \beta_c'\mu_x\right)^2.
\end{aligned}
$$

Therefore,

$$\nu_a = \frac{1}{\sigma_{e_c}^2}\left(\mu_c - \beta_c'\mu_x\right)^2.$$

Similar computations yields

$$
\begin{aligned}
b_{x_a} &= b_x + \frac{1}{\sigma_{e_c}^2}\left(\mu_c - \beta_c'\mu_x\right)\left(1 - \beta_c'\ell_N\right) \\
c_{x_a} &= c_x + \frac{1}{\sigma_{e_c}^2}\left(1 - \beta_c'\ell_N\right)^2.
\end{aligned}
$$

Consequently,

$$\nu_b = \frac{1}{\sigma_{e_c}^2}\left(\mu_c - \beta_c'\mu_x\right)\left(1 - \beta_c'\ell_N\right) \text{ and } \nu_c = \frac{1}{\sigma_{e_c}^2}\left(1 - \beta_c'\ell_N\right)^2,$$

and thus $\nu_b^2 = \nu_a\nu_c$. Plugging these solutions to (36) leads to Equation (14). Then, it is straightforward to show that the minimum value of the difference in the squared second-norm is zero at $\mu_m = \frac{1-\beta_c'\ell_N}{\mu_c - \beta_c'\mu_x}$. *q.e.d.*

Proof of Corollary 1: $\|(m_{x_a}^*|\mu_m)^2\| - \|(m_x^*|\mu_m)^2\| = 0$ for all μ_m if and only if $(\alpha_{c|\mu_m}^*) = 0$ for all μ_m. Given $\alpha_{c|\mu_m}^* = (\mu_c - \beta_c'\mu_x)\mu_m - (1 - \beta_c'\ell_N)$, the two conditions should be satisfied simultaneously. *q.e.d.*

Proof of Proposition 2: Since $m_{x_a|\mu_m} \in \mathscr{M}(x_a) \subset \mathscr{M}(x)$, there exists $\varepsilon_{c|\mu_m}$, which orthogonally extends $m^*_{x|\mu_m}$ to $m_{x_a|\mu_m} \in \mathscr{M}(x_a)$ while keeping the same mean such that:

$$1 = E[m_{x_a|\mu_m} \cdot x_c] = E[(m^*_{x|\mu_m} + \varepsilon_{c|\mu_m}) \cdot x_c] = E[m^*_{x|\mu_m} \cdot x_c] + E[\varepsilon_{c|\mu_m} \cdot x_c]. \tag{37}$$

From the projection theorem, the first term in the right-hand-side of Equation (37) is

$$\begin{aligned} E[m^*_{x|\mu_m} \cdot x_c] &= E[m^*_{x|\mu_m} \cdot (\beta_{c0} + \beta_c' x + e_c)] \\ &= \mu_m \beta_{c0} + \beta_c' E[m^*_{x|\mu_m} \cdot x] + E[m^*_{x|\mu_m} \cdot e_c] \\ &= \mu_m \beta_{c0} + \beta_c' \ell_N \\ &= (\mu_c - \beta_c' \mu_x)\mu_m + \beta_c' \ell_N. \end{aligned} \tag{38}$$

In the above, $E[m^*_{x|\mu_m} \cdot e_c] = 0$ since $e_c \perp m^*_{x|\mu_m}$. Therefore,

$$E[m^*_{x|\mu_m} \cdot x_c] - 1 = (\mu_c - \beta_c' \mu_x)\mu_m + \beta_c' \ell_N - 1 = \alpha^*_{c|\mu_m},$$

which is the proof of Equation (18). In addition, the second term in (37) is

$$\begin{aligned} E[\varepsilon_{c|\mu_m} \cdot x_c] &= E[\varepsilon_{c|\mu_m} \cdot (\beta_{c0} + \beta_c' x + e_c)] \\ &= \beta_{c0} E[\varepsilon_{c|\mu_m}] + \beta_c' E[\varepsilon_{c|\mu_m} \cdot x] + E[\varepsilon_{c|\mu_m} \cdot e_c] \\ &= E[\varepsilon_{c|\mu_m} \cdot e_c]. \end{aligned}$$

Note that $E[\varepsilon_{c|\mu_m}] = 0$ since $\varepsilon_{c|\mu_m}$ should not change the mean of the SDF. In addition, $E[\varepsilon_{c|\mu_m} \cdot x] = 0_N$ since $\varepsilon_{c|\mu_m} \in \mathscr{A}(x)^{\perp}$. Therefore, Equation (37) can be rewritten as:

$$-\alpha^*_{c|\mu_m} = E(\varepsilon_{c|\mu_m} \cdot e_c), \tag{39}$$

which is the proof of Equation (19). From the Cauchy-Schwarz inequality,

$$|-\alpha^*_{c|\mu_m}| = |E(\varepsilon_{c|\mu_m} \cdot e_c)| \leq \|\varepsilon_{c|\mu_m}\| \cdot \|e_c\|. \tag{40}$$

In addition, $\mathscr{M}(x_a) \cap \mathscr{A}((1/\mu_m\ x_a')') \neq \emptyset$ and thus the Riesz Representation Theorem dictates the existence of a marketd payoff, $\varepsilon^*_{c|\mu_m} = \lambda_0 \frac{1}{\mu_m} + \lambda_x' x + \lambda_c x_c$ where $\lambda' = (\lambda_0\ \ \lambda_x'\ \ \lambda_c) \in \mathbb{R}^{N+2}$.[29] Since the extended SDF should maintain the same mean,

$$E(\varepsilon^*_{c|\mu_m}) = \lambda_0 \frac{1}{\mu_m} + \lambda_x' \mu_x + \lambda_c \mu_c = 0,$$

which yields $\lambda_0 = -\mu_m \cdot (\lambda_x' \mu_x + \lambda_c \mu_c)$. Therefore the matketed extension of the SDF is simplified to

$$\varepsilon^*_{c|\mu_m} = \lambda_x'(x - \mu_x) + \lambda_c(x_c - \mu_c).$$

Then, given the orthogonality between x and e_c,

$$\begin{aligned} E(\varepsilon^*_{c|\mu_m} \cdot e_c) &= E[(\lambda_x'(x - \mu_x) + \lambda_c(x_c - \mu_c)) \cdot e_c] \\ &= \lambda_x' E[(x - \mu_x) \cdot e_c] + \lambda_c E[(\beta_{c0} + \beta' x + e - \mu_c) \cdot e_c] \\ &= \lambda_c \sigma^2_{e_c}. \end{aligned}$$

[29] Given μ_m, we implicitly assumes the existence of the risk-free asset with the gross risk-free interest rate, $x_f = 1/\mu_m$. Then the attainable set is spanned from the augmented set of assets which are composed of the implicit risk-free asset (x_f), the stock basis assets (x) and the cryptoasset (x_c).

Combining this result with (39) leads to

$$\lambda_c = \frac{-\alpha^*_{c|\mu_m}}{\|e\|^2}.$$

Finally, we can determine λ_x from $E[\varepsilon^*_{c|\mu_m} \cdot x] = 0_N$.

$$\begin{aligned}
0_N &= E(\varepsilon^*_{c|\mu_m} \cdot x) \\
&= E\left[\left(\lambda'_x(x-\mu_x) + \lambda_c(x_c - \mu_c)\right) \cdot x\right] \\
&= E[(x-\mu_x)x']\lambda_x + \lambda_c E[(\beta_{c0} + \beta'_c x + e_c - \mu_c) \cdot x] \\
&= \Sigma_x \lambda_x + \lambda_c[(\beta_{c0} - \mu_c)\mu_x + E(xx')\beta] \\
&= \Sigma_x \lambda_x + \lambda_c[-\mu_x \mu'_x \beta_c + E(xx')\beta_c] \\
&= \Sigma_x(\lambda_x + \lambda_c \beta_c),
\end{aligned}$$

where we use $\beta_{c0} = \mu_c - \beta'_c \mu_x$. Therefore, $\lambda_x = -\lambda_c \beta_c = \frac{\alpha^*_{c|\mu_m}}{\|e\|^2}\beta_c$. Combining all these results, we end up with

$$\begin{aligned}
\varepsilon^*_{c|\mu_m} &= \lambda'_x(x-\mu_x) + \lambda_c(x_c - \mu_c) \\
&= \lambda_c\left[-\beta'_c(x-\mu_x) + (x_c - \mu_c)\right] \\
&= \lambda_c \cdot e_c \\
&= \frac{-\alpha^*_{c|\mu_m}}{\|e_c\|^2} \; e_c.
\end{aligned}$$

Therefore, the second-norm of $\varepsilon^*_{c|\mu_m}$ is

$$\|\varepsilon^*_{c|\mu_m}\| = \frac{|-\alpha^*_{c|\mu_m}|}{\|e_c\|^2} \cdot \|e_c\| = \frac{|\alpha^*_{c|\mu_n}|}{\|e_c\|},$$

and its square is the same to the right-hand-side of Equation (14) and thus $\|\varepsilon^*_{c|\mu_m}\|^2 = \|m^*_{x_a|\mu_m}\| - \|m^*_{x|\mu_m}\|$. Finally, we prove that the marketd extension, $\varepsilon^*_{c|\mu_m}$, has the minimum second-norm among $\varepsilon_{c|\mu_m}$s by showing that Equation (40) holds with equality.

$$\|\varepsilon^*_{c|\mu_m}\| \cdot \|e_c\| = \frac{|\alpha^*_{c|\mu_m}|}{\|e_c\|} \cdot \|e_c\| = |\alpha^*_{c|\mu_m}|,$$

and therefore Equation (40) holds with equality. *q.e.d.*

Proof of Proposition 3: We prove (i) and (ii) in a sequence.

(i) From Equation (22), $\|\varepsilon^*_{c|\mu_m}\| = 0$ if and only if $\alpha^*_{c|m} = 0$. Corollary 1 proves that $\alpha^*_{c|\mu_m} = 0$ for all μ_m except $\overline{\mu}_m$ if and only if the two conditions specified in (16) are met: i.e., $\beta'_c \mu_x = \mu_c$ and $\beta'_c \ell_N = 1$. If the two conditions are 'perfectly' satisfied, $\overline{\mu}_m = \frac{0}{0}$ is not defined and this is concrete evidence that the frontier does not move at all. However, in a sample, these two conditions hold $\mathscr{P}$-*almost never*. Therefore, the estimate of $\overline{\mu}_m$ is defined $\mathscr{P}$-*almost surely* in a sample sapce. Combining all of these results, $\alpha^*_{c|\mu} \neq 0$ (or equivalently $\|\varepsilon^*_{c|\mu_m}\| > 0$) for any arbitrary μ_m other than $\overline{\mu}_m$ implies that these two conditions are not satisfied and thus

$\|\varepsilon^*_{c|\mu_n}\| > 0$ for all μ_n other than the estimate of $\overline{\mu}_m$ as well. Geometrically this means an upward shift in the mean-second norm frontier of the SDF. In sum, $\alpha^*_{c|\mu} \neq 0$ (or equivalently $\|\varepsilon^*_{c|\mu_m}\| > 0$) for any arbitrary μ_m other than the estimate of $\overline{\mu}_m$ is equivalent to the condition of the crypto being neither replicable nor aligned in price with the stock basis assets.

(ii) Suppose that the cryptoasset makes an upward shift of the frontier. However, even in such a case, there still exists a unique solution, $\overline{\mu}_m = \frac{1-\beta_c'\ell_N}{\mu_c-\beta_c'\mu_x}$, which makes $\alpha^*_{c|\mu_m} = 0$ or equivalently $\|\varepsilon^*_{c|x}\| = 0$ as shown in Proposition 1. Equation (18) in Proposition 2 can be rewritten in the form of the expected gross return:

$$\frac{(\alpha^*_{c|\mu_m})}{\mu_m} = (\mu_c - \beta_c'\mu_x) - \frac{1}{\mu_m}(1 - \beta_c'\ell_N). \tag{41}$$

Let $\overline{\mu}_f$ denote the inverse of $\overline{\mu}_m$, i.e., $\overline{\mu}_f = \frac{1}{\overline{\mu}_m} = \frac{\mu_c-\beta_c'\mu_x}{1-\beta_c'\ell_N}$. Plugging $\overline{\mu}_f$ into Equation (41) results in

$$0 = \mu_c - \left[(1 - \beta_c'\ell_N)\overline{\mu}_f + \beta_c'\mu_x\right]. \tag{42}$$

Therefore, $\overline{\mu}_f$ is the required level of the risk-free rate which equates the expected gross return on the crypto to that on the replicating portfolio. Unlike in (i), Equation (42) shows that the replicating portfolio herein includes the risk-free asset; the investment weight on it is $1-\beta_c'\ell_N$ whereas the weights on the stock basis assets are β_c. Put differently, $\overline{\mu}_f$ can be thought of a shadow interest rate implied by the cryptoasset, which is a solution to Equation (42). In the presence of the market risk-free asset, the risk-free rate should be the same to this shadow risk-free rate. If so, the cryptoasset is replicable by and is aligned in price with the augmented set of basis assets (the stock basis assets+the risk-free asset) and vice versa.

This completes the proof. *q.e.d.*

Proof of Proposition 4: Because $m^*_{\mathbb{e}}$ itself is the GMN-SDF of the extended set of basis assets, we can apply Equation (12) to $m^*_{\mathbb{e}}$, which yields $\mu^*_{m_{\mathbb{e}}} = \mu'_{x_{\mathbb{e}}}E(x_{\mathbb{e}}x'_{\mathbb{e}})^{-1}\ell_{N+1}$. The Sherman-Morrison formula leads to

$$\begin{aligned}\mu^*_{m_{\mathbb{e}}} &= \mu'_{x_{\mathbb{e}}}\left(\Sigma^{-1}_{x_{\mathbb{e}}} - \frac{\Sigma^{-1}_{x_{\mathbb{e}}}\mu_{x_{\mathbb{e}}}\mu'_{x_{\mathbb{e}}}\Sigma^{-1}_{x_{\mathbb{e}}}}{1+\mu'_{x_{\mathbb{e}}}\Sigma^{-1}_{x_{\mathbb{e}}}\mu_{x_{\mathbb{e}}}}\right)\ell_{N+1} \\ &= \frac{\mu'_{x_{\mathbb{e}}}\Sigma^{-1}_{x_{\mathbb{e}}}\ell_{N+1}}{1+\mu'_{x_{\mathbb{e}}}\Sigma^{-1}_{x_{\mathbb{e}}}\mu_{x_{\mathbb{e}}}}.\end{aligned} \tag{43}$$

Because $x_f \perp x$,

$$\Sigma^{-1}_{x_{\mathbb{e}}} = \begin{bmatrix} \frac{1}{\sigma_f^2} & 0'_{N+1} \\ 0_{N+1} & \Sigma_x^{-1} \end{bmatrix}.$$

Therefore, $\mu'_{x_{\mathbb{e}}}\Sigma^{-1}_{x_{\mathbb{e}}}\mu_{x_{\mathbb{e}}} = \frac{\mu_f^2}{\sigma_f^2} + \mu'_x\Sigma_x^{-1}\mu_x = \frac{\mu_f^2}{\sigma_f^2} + a_x$ and $\mu'_{x_{\mathbb{e}}}\Sigma^{-1}_{x_{\mathbb{e}}}\ell_{N+1} = \frac{\mu_f^2}{\sigma_f^2} + \mu'_x\Sigma_x^{-1}\ell_N = \frac{\mu_f^2}{\sigma_f^2} + b_x$. Plugging these results into (43) yields the first equation of (26). Applying the first-order Taylor expansion around $\sigma_f^2 = 0$ yields the second equation. In addition, Equation (15) can be transformed into:

$$
\begin{aligned}
\alpha^*_{c|\mu^*_{m_{\mathbb{e}}}} &= (\mu_c - \beta'_{c_{\mathbb{e}}}\mu_{x_{\mathbb{e}}})\mu^*_{m_{\mathbb{e}}} - (1 - \beta'_{c_{\mathbb{e}}}\ell_{N+1}) \\
&= (\mu_c - \beta_{c_f}\mu_f - \beta'_c\mu_x)\mu^*_{m_{\mathbb{e}}} - (1 - \beta_{c_f} - \beta'_c\ell_N) \\
&= \beta_{c_f}(1 - \mu_f\mu^*_{m_{\mathbb{e}}}) + (\mu_c - \beta'_c\mu_x)\mu^*_{m_{\mathbb{e}}} - (1 - \beta'_c\ell_N). \qquad (44)
\end{aligned}
$$

where β_{c_f} is the investment weight on the locally risk-free asset. In the first term of the last equation, $\mu_f\mu^*_{m_{\mathbb{e}}}$ is

$$
\mu_f\mu^*_{m_{\mathbb{e}}} = \mu_f\left(\frac{\mu_f + b_x\sigma_f^2}{\mu_f^2 + (1+a_x)\sigma_f^2}\right) \approx \mu_f\left(\frac{1}{\mu_f} + \upsilon_J\right) = 1 + \mu_f\upsilon_J.
$$

Pluggin this result back into Equation (44) yields

$$
\alpha^*_{c|\mu^*_{m_{\mathbb{e}}}} \approx (\mu_c - \beta'_c\mu_x)\frac{1}{\mu_f} - (1 - \beta'_c\ell_N) + (\mu_c - \beta'_c\mu_x - \beta_{c_f}\mu_f)\upsilon_J,
$$

and $\upsilon_J \approx 0$ yields Equation (27).

Lemma A.1: *The pricing error of x_c against the GMN-SDF is*

$$
\alpha^*_c = (\beta'_c\ell_N - 1) - (\beta'_c\mu_x - \mu_c)\frac{b_x}{1+a_x} = \mathscr{b}'_c\ell_N - 1. \qquad (45)
$$

Proof: From the Sherman-Morrison formula, we get

$$
\Sigma_x^{-1} = E(xx')^{-1} + \frac{\Sigma_x^{-1}\mu_x\mu'_x\Sigma_x^{-1}}{1 + \mu'_x\Sigma_x^{-1}\mu_x}.
$$

Using this result and the definition of β_c,

$$
\begin{aligned}
\beta'_c\ell_N - 1 &= \Sigma'_{xc}\Sigma_x^{-1}\ell_N - 1 \\
&= \big(E(x_c x) - \mu_c\mu_x\big)'\Sigma_x^{-1}\ell_N - 1 \\
&= E(x_c x')\left[E(xx')^{-1} + \frac{\Sigma_x^{-1}\mu_x\mu'_x\Sigma_x^{-1}}{1 + \mu'_x\Sigma_x^{-1}\mu_x}\right]\ell_N - \mu_c(\mu'_x\Sigma_x^{-1}\ell_N) - 1 \\
&= E(x_c x')E(xx')^{-1}\ell_N + \frac{E(x_c x')\Sigma_x^{-1}\mu_x}{1+a_x}b_x - \mu_c b_x - 1 \\
&= \mathscr{b}'_c\ell_N + \frac{(\Sigma_{xc} + \mu_c\mu_x)'\Sigma_x^{-1}\mu_x}{1+a_x}b_x - \mu_c b_x - 1 \\
&= \mathscr{b}'_c\ell_N + \frac{\beta'_c\mu_x + \mu_c a_x}{1+a_x}b_x - \mu_c b_x - 1 \\
&= \mathscr{b}'_c\ell_N - 1 + \frac{\beta'_c\mu_x - \mu_c}{1+a_x}b_x,
\end{aligned}
$$

and this completes the proof of Equation (45). *q.e.d.*

Lemma A.2: *The squared second-norm of ω is*

$$
\|\omega_c\|^2 = \sigma^2_{e_c} + (1+a_x)E(\omega_c)^2, \qquad (46)
$$

and

$$E(\omega_c) = \frac{\beta_{c0}}{1 + a_x}, \tag{47}$$

and thus the alternative expression of $\|\omega_c\|^2$ *is*

$$\|\omega_c\|^2 = \sigma_{e_c}^2 + \frac{\beta_{c0}^2}{1 + a_x}. \tag{48}$$

Proof: The two alternative idiosyncratic components are

$$\begin{aligned} e_c &= (x_c - \mu_c) - \beta_c'(x - \mu_x) \\ \omega_c &= x_c - b_c' x, \end{aligned}$$

and thus $e_c = \omega_c - (\mu_c - \beta_c'\mu_x) - (\beta_c - b_c)'x$, which yields

$$\sigma_{e_c}^2 = E(\omega_c^2) - E(\omega_c)^2 - (\beta_c - b_c)'\Sigma_x(\beta_c - b_c) \implies E(\omega_c^2) = \sigma_{e_c}^2 + E(\omega_c)^2 + (\beta_c - b_c)'\Sigma_x(\beta_c - b_c), \tag{49}$$

where

$$\begin{aligned} \beta_c - b_c &= \Sigma_x^{-1}\Sigma_{xc} - E(xx')^{-1}E(xx_c) \\ &= \Sigma_x^{-1}\big(E(xx_c) - \mu_x\mu_c\big) - \left[\Sigma_x^{-1} - \frac{\Sigma_x^{-1}\mu_x\mu_x'\Sigma_x^{-1}}{1 + a_x}\right] E(xx_c) \\ &= \frac{\Sigma_x^{-1}\mu_x\mu_x'\Sigma_x^{-1}}{1 + a_x} E(xx_c) - \Sigma_x^{-1}\mu_x\mu_c. \end{aligned}$$

Therefore,

$$\begin{aligned} (\beta_c - b_c)'\Sigma_x(\beta_c - b_c) &= \frac{E(x_c x')\Sigma_x^{-1}\mu_x\mu_x'\Sigma_x^{-1}}{1 + a_x} \cdot \Sigma_x \cdot \frac{\Sigma_x^{-1}\mu_x\mu_x'\Sigma_x^{-1}E(xx_c)}{1 + a_x} \\ &\quad -2\frac{E(x_c x')\Sigma_x^{-1}\mu_x\mu_x'\Sigma_x^{-1}}{1 + a_x} \cdot \Sigma_x \cdot \Sigma_x^{-1}\mu_x\mu_c \\ &\quad +\mu_c^2\mu_x'\Sigma_x^{-1} \cdot \Sigma_x \cdot \Sigma_x^{-1}\mu_x \\ &= \frac{a_x}{(1 + a_x)^2}\left[E(x_c x')\Sigma_x^{-1}\mu_x\right]^2 - 2\frac{a_x}{1 + a_x}\left[E(x_c x')\Sigma_x^{-1}\mu_x\right]\mu_c + a_x\mu_c^2 \\ &= a_x\left[\frac{E(x_c x')\Sigma_x^{-1}\mu_x}{1 + a_x} - \mu_c\right]^2. \end{aligned} \tag{50}$$

Therein,

$$\begin{aligned} E(x_c x')\Sigma_x^{-1}\mu_x &= E(x_c x')\left[E(xx')^{-1} + \frac{\Sigma_x^{-1}\mu_x\mu_x'\Sigma_x^{-1}}{1 + a_x}\right]\mu_x \\ &= b_c'\mu_x + \frac{E(x_c x')\Sigma_x^{-1}\mu_x \; a_x}{1 + a_x}, \end{aligned}$$

and we can solve this equation for $E(x_c x')\Sigma_x^{-1}\mu_x$:

$$E(x_c x')\Sigma_x^{-1}\mu_x = (1 + a_x)b_c'\mu_x. \tag{51}$$

Plugging this into (50) yields

$$\begin{aligned}(\beta_c - b_c)'\Sigma_x(\beta_c - b_c) &= a_x(b_c'\mu_x - \mu_c)^2 \\ &= a_x E(\omega_c)^2,\end{aligned}$$

and thus (49) can be rewritten as $E(\omega^2) = \sigma_{e_c}^2 + (1 + a_x)E(\omega)^2$, which is Equation (46). Further we can re-solve $E(x_c x')\Sigma_x^{-1}\mu_x$ in the following way:

$$\begin{aligned}E(x_c x')\Sigma_x^{-1}\mu_x &= \left[E(x_c x') - \mu_c\mu_x'\right]\Sigma_x^{-1}\mu_x + \mu_c\mu_x'\Sigma_x^{-1}\mu_x \\ &= \beta_c'\mu_x + a_x\mu_c. \end{aligned} \tag{52}$$

Equating (52) and (51), we get $b_c'\mu_x = \frac{\beta_c'\mu_x + a_x\mu_c}{1+a_x}$, which yields

$$E(\omega_c) = \mu_c - b_c'\mu_x = \mu_c - \frac{\beta_c'\mu_x + a_x\mu_c}{1 + a_x} = \frac{\mu_c - \beta_c'\mu_x}{1 + a_x} = \frac{\beta_{c0}}{1 + a_x},$$

which is Equation (47). Plugging this result into (46) yields (48). *q.e.d.*

Proof of Proposition 5: The proof is similar to that of Proposition 2. Since $m_{x_a} \in \mathscr{M}(x_a) \subset \mathscr{M}(x)$, there exists ε_c, which orthogonally extends m_x^* to any $m_{x_a} \in \mathscr{M}(x_a)$ such that:

$$1 = E[m_{x_a} \cdot x_c] = E[(m_x^* + \varepsilon_c) \cdot x_c] = E[m_x^* \cdot x_c] + E[\varepsilon_c \cdot x_c]. \tag{53}$$

We project x_c on x to construct a mimicking portfolio of x_c

$$x_c = b_c'x + \omega_c; \quad \text{where } \omega_c \perp x,$$

and the projection theorem leads to $b = E(xx')^{-1}E(xx_c)$. Then, the first term in (53) is

$$\begin{aligned}E[m_x^* \cdot x_c] &= E[m_x^*(b_c'x + \omega_c)] \\ &= b_c'E(m_x^* \cdot x) + E(m_x^* \cdot \omega_c) \\ &= b_c'\ell_N.\end{aligned}$$

In the above, $E(m_x^* \cdot \omega_c) = 0$ since m_x^* is a linear function of x to which ω_c is orthogonal. In addition, the second term in (53) is

$$\begin{aligned}E(\varepsilon_c \cdot x_c) &= E[\varepsilon_c(b_c'x + \omega_c)] \\ &= b_c'E(\varepsilon_c \cdot x) + E(\varepsilon_c \cdot \omega) \\ &= E(\varepsilon_c \cdot \omega_c).\end{aligned}$$

Note that $E(\varepsilon_c \cdot x) = 0$ since $\varepsilon_c \in \mathscr{A}(x)^\perp$. Therefore, Equation (53) can be rewritten as:

$$1 - b_c'\ell_N = E(\varepsilon_c \cdot \omega_c). \tag{54}$$

Note that $1 - b_c'\ell_N = -\alpha_c^*$. From the Cauchy-Schwarz inequality,

$$\left|\alpha_c^*\right| = \left|1 - b_c'\ell_N\right| = |E(\varepsilon_c \cdot \omega)| \leq \|\varepsilon_c\|\|\omega_c\|. \tag{55}$$

In addition, $\mathscr{M}(x_a) \cap \mathscr{A})(x_a)$ and thus the Riesz Representation Theorem dictates the existence of $\varepsilon_c^* = \psi' x_a$ where $\psi' = (\psi_x' \;\; \psi_c) \in \mathbb{R}^{N+1}$ Then, given the orthogonality between x and ω,

$$\begin{aligned} E(\varepsilon_c^* \cdot \omega_c) &= E[(\psi_x' x + \psi_c x_c) \cdot \omega_c] \\ &= E[(\psi_x' x + \psi_c(\mathscr{b}_c' x + \omega_c)) \cdot \omega_c] \\ &= E[(\psi_x + \psi_c \mathscr{b}')x \cdot \omega_c] + \psi_c E(\omega_c^2) \\ &= \psi_c E(\omega_c^2). \end{aligned}$$

Therefore,

$$\psi_c = \frac{1 - \mathscr{b}' \ell_N}{\|\omega_c\|^2},$$

which is a sort of 'normalized mispricing.' We can determine ψ_x from $E(\varepsilon_c^* \cdot x) = 0_N$.

$$\begin{aligned} 0_N &= E(\varepsilon_c^* \cdot x) \\ &= E[(\psi_x' x + \psi_c x_c) \cdot x] \\ &= E[(\psi_x' x + \psi_c(\mathscr{b}_c' x)) \cdot x] \\ &= E(xx')(\psi_x + \psi_c \mathscr{b}_c), \end{aligned}$$

which yields $\psi_x = -\psi_c \mathscr{b}_c = -\frac{(1-\mathscr{b}_c'\ell_N))\mathscr{b}_c}{\|\omega_c\|^2}$. Therefore,

$$\begin{aligned} \varepsilon_c^* &= \psi_x' x + \psi_c x_c \\ &= -\frac{(1 - \mathscr{b}_c' \ell_N)\, \mathscr{b}_c'}{\|\omega_c\|^2} x + \frac{1 - \mathscr{b}_c' \ell_N}{\|\omega_c\|^2} x_c \\ &= \frac{1 - \mathscr{b}_c' \ell_N}{\|\omega_c\|^2} (x_c - \mathscr{b}_c' x) \\ &= \frac{1 - \mathscr{b}_c' \ell_N}{\|\omega_c\|^2} \omega_c \\ &\overset{or}{=} \psi_c \omega_c \\ &\overset{or}{=} \frac{-\alpha_c^*}{\|\omega_c\|} \cdot \omega_c. \end{aligned}$$

Therefore, the second-norm of ε^* is

$$\|\varepsilon_c^*\| = \frac{|1 - \mathscr{b}_c' \ell_N|}{\|\omega_c\|^2} \cdot \|\omega_c\| = \frac{|\alpha_c^*|}{\|\omega_c\|}.$$

We prove that the second-norm of ε^* is, in fact, the minimum distance from m_x^* to $\mathscr{M}(x_a)$ by showing that (55) holds with equality:

$$\|\varepsilon_c^*\| \cdot \|\omega_c\| = \frac{|\alpha_c^*|}{\|\omega_c\|} \cdot \|\omega_c\| = |\alpha_c^*| = |E(\varepsilon_c^* \cdot \omega_c)|.$$

Finally, we prove that $\|\varepsilon_c^*\| = \|m_{x_a}^*\| - \|m_x^*\|$. From Equation (12) and the definitions of ν_a, ν_b and ν_c in the proof of Proposition 1:

$$\begin{aligned}
\|m_{x_a}^*\|^2 - \|m_x^*\|^2 &= \frac{(1+a_{x_a})c_{x_a} - b_{x_a}^2}{(1+a_{x_a})} - \frac{(1+a_x)c_x - b_x^2}{(1+a_x)} \\
&= \frac{(1+a_x)(\nu_a\nu_c - \nu_b^2) + (1+a_x)^2\nu_c + b_x^2\nu_a - 2(1+a_x)b_x\nu_b}{(1+a_x)(1+a_x+\nu_a)} \\
&= \frac{(1+a_x)^2\nu_c + b_x^2\nu_a - 2(1+a_x)b_x\nu_b}{(1+a_x)(1+a_x+\nu_a)} \\
&= \frac{(1+a_x)^2\,\frac{(1-\beta_c'\ell_N)^2}{\sigma_{e_c}^2} + b_x^2\,\frac{(\mu_c-\beta_c'\mu_x)^2}{\sigma_{e_c}^2} - 2(1+a_x)b_x\frac{(1-\beta_c'\ell_N)(\mu_c-\beta_c'\mu_x)}{\sigma_{e_c}^2}}{(1+a_x)(1+a_x+\nu_a)} \\
&= \frac{\left[(1+a_x)(1-\beta_c'\ell_N) - b_x(\mu_c-\beta_c'\mu_x)\right]^2}{\sigma_{e_c}^2(1+a_x)(1+a_x+\nu_a)} \\
&= \frac{(1+a_x)\left[(1-\beta_c'\ell_N) - \frac{b_x}{1+a_x}(\mu_x-\beta_c'\mu_x)\right]^2}{\sigma_{e_c}^2(1+a_x+\nu_a)} \\
&= \frac{(1+a_x)(-\alpha_c^*)^2}{\sigma_{e_c}^2(1+a_x+\nu_a)} \\
&= \frac{\alpha_c^{*2}}{\sigma_{e_c}^2\left(1+\frac{\nu_a}{1+a_x}\right)}.
\end{aligned} \tag{56}$$

In the above, we used the fact that $\nu_a\nu_c = \nu_b^2$. To complete the proof, we need to show that $\alpha_c^* = \beta_c'\ell_N - 1$ and $\sigma_{e_c}^2\left(1+\frac{\nu_a}{1+a_x}\right) = E(\omega_c^2)$. Lemma A.1 proves the first, $\alpha_c^* = \beta_c'\ell_N - 1$. The second one is

$$\begin{aligned}
\sigma_{e_c}^2\left(1+\frac{\nu_a}{1+a_x}\right) &= \sigma_{e_c}^2\left[1+\frac{(\mu_c-\beta_c'\mu_x)^2}{(1+a_x)\sigma_{e_c}^2}\right] \\
&= \sigma_{e_c}^2 + \frac{\beta_{c0}^2}{1+a_x} \\
&= E(\omega_c^2),
\end{aligned}$$

where the last equality comes from Lemma A.2. This completes the proof. *q.e.d.*

Proof of Corollary 2: Given Proposition 5, proofs are straightforward. First, if $\|\varepsilon_c^*\| \neq 0$, which means $\alpha_c^* \neq 0$. From Corollary 1, this means that both of two conditions, $\mu_c = \beta'\mu_x$ and $\beta'\ell_N = 1$, are not simultaneously satisfied. In such a case, Proposition 3 states that the mean-second norm frontier of the SDF shifts up. The opposite is case (i) wherein both conditions are met. Then, the frontier retrieved from the augmented set of assets precisely coincides with that retrieved from the risky basis assets. Finally, case (ii) occurs when the tangent point derived in Proposition 1 occurs coincidently at the GMN-SDF. As such, the frontier overall shifts up but the two frontiers meet each other with tangency at the minimum norm. At other points, the frontier retrieved from the augmented set of assets is located above. However, $\alpha_c^* = \beta_c'\ell_N - 1$ holds only at μ_m^*. At other μ_m, this relation does not hold and the corresponding non-zero $\alpha_{c|\mu_m}^*$ is based on the regression (17), which includes the intercept term or eqivalently the risk-free asset. However, it violates the assumption that the risk-free asset does not exist. As such, the shift-up of the frontier at the mean value other than μ_m^* does not deliver any economic meaning. All in all case (ii) is not different from case (i) in its economic implication.

Lemma A.3: *The followings are true:*

$$\begin{aligned}
\mu_x' E(xx')^1 \mu_x &= \frac{a_x}{1+a_x} \\
\ell_N' E(xx')^{-1} \ell_N &= \frac{c_x + c_x a_x - b_x^2}{1+a_x} \\
\mu_x' E(xx')^{-1} \ell_N &= \frac{b_x}{1+a_x}.
\end{aligned}$$

Proof: From Lemma A.1., the Sherman-Morrison formula delivers

$$E(xx')^{-1} = \Sigma_x^{-1} - \frac{\Sigma_x^{-1}\mu_x\mu_x'\Sigma_x^{-1}}{1+\mu_x'\Sigma_x^{-1}\mu_x} = \frac{1}{1+\mu_x'\Sigma_x^{-1}\mu_x}\left[\Sigma_x^{-1} + \Sigma_x^{-1}\left(\mu_x'\Sigma_x^{-1}\mu_x\right) - \Sigma_x^{-1}\mu_x\mu_x'\Sigma_x^{-1}\right].$$

Then,

$$\begin{aligned}
\mu_x' E(xx')^{-1} \mu_x &= \frac{1}{1+a_x}\left[\mu_x'\Sigma_x^{-1}\mu_x + \left(\mu_x'\Sigma_x^{-1}\mu_x\right)^2 - \mu_x'\Sigma_x^{-1}\mu_x\mu_x'\Sigma_x^{-1}\mu_x\right] \\
&= \frac{a_x}{1+a_x} \\
\mu_x' E(xx')^{-1} \ell_N &= \frac{1}{1+a_x}\left[\mu_x'\Sigma_x^{-1}\ell_N + \left(\mu_x'\Sigma_x^{-1}\ell_N\right)\left(\mu_x'\Sigma_x^{-1}\mu_x\right) - \mu_x'\Sigma_x^{-1}\mu_x\mu_x'\Sigma_x^{-1}\ell_N\right] \\
&= \frac{b_x}{1+a_x} \\
\ell_N' E(xx')^{-1} \ell_N &= \frac{1}{1+a_x}\left[\ell_N'\Sigma_x^{-1}\ell_N + \left(\ell_N'\Sigma_x^{-1}\ell_N\right)\left(\mu_x'\Sigma_x^{-1}\mu_x\right) - \ell_N'\Sigma_x^{-1}\mu_x\mu_x'\Sigma_x^{-1}\ell_N\right] \\
&= \frac{c_x + c_x a_x - b_x^2}{1+a_x}.
\end{aligned}$$

Recall that $\mu_x'\Sigma_x^{-1}\mu_x = a_x$, $\mu_x'\Sigma_x^{-1}\ell_N = b_x$ and $\ell_N'\Sigma_x^{-1}\ell_N = c_x$. *q.e.d.*

Proof of Proposition 6: We derive the exact sample distribution of $\hat{\alpha}_c^*$. From Lemma A.1, we know that $\alpha_c^* = (\beta_c'\ell_N - 1) - (\beta_c'\mu_x - \mu_c)\frac{b_x}{1+a_x} = \mathscr{b}_c'\ell_N - 1$. Herein we derive the sample distribution based on $\hat{\mathscr{b}}_c{}'\ell_N - 1$. The OLS estimator of $\mathscr{b}_c$ is

$$\hat{\mathscr{b}}_c = (X'X)^{-1}X'X_c = (X'X)^{-1}X'(\beta_{c0}\ell_T + X\beta_c + \mathbf{e_c}) = \beta_c + \beta_{c0}(X'X)^{-1}X'\ell_T + (X'X)^{-1}X'\mathbf{e_c},$$

where X and X_c denote a $T \times N$ sample matrix and $T \times 1$ sample vector corresponding x and x_c respectively. Similarly $\mathbf{e_c}$ is $T \times 1$ vector of **true** residuals. We assume $\mathbf{e_c} \sim \mathcal{N}(0_T, \sigma_{e_c}^2 I_T)$, i.e., is normally distributed with zero mean and variance of $\sigma_{e_c}^2 I_T$. Therefore,

$$\hat{\alpha}_c^* = \hat{\mathscr{b}}_c{}'\ell_N - 1 = \beta_c'\ell_N + \beta_{c0}\ell_T'X(X'X)^{-1}\ell_N + \mathbf{e_c}'X(X'X)^{-1}\ell_N - 1, \tag{57}$$

where

$$\ell_T'X(X'X)^{-1}\ell_N = \left(\frac{\ell_T'X}{T}\right)\left(\frac{X'X}{T}\right)^{-1}\ell_N = \hat{\mu}_x'\hat{\Omega}_x^{-1}\ell_N,$$

where $\hat{\mu}_x = \frac{X'\ell_T}{T} = \frac{1}{T}\sum_{t=1}^{T} X_t$ is the estimate of μ_x and similarly $\hat{\Omega}_x = \frac{X'X}{T} = \frac{1}{T}\sum_{t=1}^{T} X_t' X_t$ is the estimator of $E(xx')$. Then, using the result of Lemma A.3,

$$\ell_T' X(X'X)^{-1}\ell_N = \hat{\mu}_x' \hat{\Omega}_x^{-1}\ell_N = \frac{\hat{b}_x}{1+\hat{a}_x},$$

where $\hat{a}_x$ and $\hat{b}_x$ denote the sample estimates of a_x and b_x. One thing to note is that X is a matrix of exogenous variables, not random variables in the least squares estimation. This implies that any variables which are a sole function of X, although sample estimates, are not exposed to any statistical inference.[30] That is, we do not need to take into account the additional statistical inference on $\hat{\mu}_x$, $\hat{\Omega}_x(=\widehat{E(xx')})$, $\hat{\Sigma}_x$, $\hat{\Sigma}_{xc}$, $\hat{a}_x$, $\hat{b}_x$ and $\hat{c}_x$ and any relations among them such as the Sherman-Morrison formula that hold in population also hold in a sample. Combining this result with $\beta_{c0} = \mu_c - \beta_c'\mu_x$, we can rewrite Equation (57) to:

$$\hat{\alpha}_c^* = \beta_c'\ell_N - 1 + (\mu_c - \beta_c'\mu_x)\frac{\hat{b}_x}{1+\hat{a}_x} + \mathbf{e_c}' X(X'X)^{-1}\ell_N = \alpha_c^* + \mathbf{e_c}' X(X'X)^{-1}\ell_N, \tag{58}$$

and thus it is unbiased. Its variance is

$$\begin{aligned}
\text{var}\left(\hat{\alpha}_c^*\right) &= \ell_N'(X'X)^{-1}X'\text{var}(\mathbf{e_c})X(X'X)^{-1}\ell_N \\
&= \sigma_{e_c}^2 \ell_N'(X'X)^{-1}\ell_N \\
&= \sigma_{e_c}^2 \ell_N' \frac{\hat{\Omega}_x^{-1}}{T}\ell_N \\
&= \frac{\sigma_{e_c}^2}{T}\frac{\hat{c}_x + \hat{c}_x\hat{a}_x - \hat{b}_x^2}{1+\hat{a}_x},
\end{aligned}$$

wherein the last equality comes from Lemma A.3. Therefore, $\hat{\alpha}_c^*$ follows the following normal distribution in a finite sample:

$$\hat{\alpha}_c^* \sim \mathcal{N}\left[\alpha_c^*,\ \frac{\sigma_{e_c}^2}{T}\cdot\frac{\hat{c}_x + \hat{c}_x\hat{a}_x - \hat{b}_x^2}{1+\hat{a}_x},\right].$$

In sum, $\hat{\alpha}_c^*$ is unbiased and its variance is proportional to the idiosyncratic variance, $\sigma_{e_c}^2$. *q.e.d.*

Lemma A.4: *Consider the following linear and quadratic forms:*

$$\begin{aligned}
Lz &= \ell_N'(X'X)^{-1}X'z \\
z'M_e z &= z'(I_T - X_e(X_e'X_e)^{-1}X_e')z,
\end{aligned}$$

where X is a $T \times N$-dimensional matrix and $X_e = (\ell_T\ X)$ is a $T \times (N+1)$-dimensional matrix, which is a concatenation of ℓ_T and X. z is a $T \times 1$ standard normal vector. Then, $LM_e = 0_T'$ and the linear function Lz and the idempotent quadratic form, $z'M_e z$, are statistically independent.

Proof: See Greene (1990), for example, for the fact that if M_e is an idempotent matrix and $LM_e = 0_T'$, Lz and $z'M_e z$ are independent. As such, the remainder of the proof is to prove M_e is idempotent and $LM_e = 0$. It is straightforward to show M_e is both symmetric, $M_e = M_e'$ and

[30] For an illustration, consider $\text{var}(\hat{\beta}) = \sigma_e^2(X'X)^{-1}$ in a standard OLS, $y = Xb + e$ with $\text{var}(e) = \sigma_e^2$. We can take $(X'X)^{-1} = \frac{\widehat{E(xx')}^{-1}}{T} = \frac{\hat{\Omega}_x^{-1}}{T}$.

idempotent, $M_e^2 = M_e$. Now let us prove $LM_e = 0$. Let $M = I_T - X(X'X)^{-1}X'$, which is also idempotent. Then, $X_e(X_e'X_e)^{-1}X_e'$ in M_e can be rewritten as:

$$
\begin{aligned}
X_e(X_e'X_e)^{-1}X_e' &= (\ell_T \; X) \begin{pmatrix} T & \ell_T'X \\ X'\ell_T & X'X \end{pmatrix}^{-1} \begin{pmatrix} \ell_T \\ X' \end{pmatrix} \\
&= (\ell_T \; X) \begin{pmatrix} \frac{1}{\ell_T'M\ell_T} & -\frac{\ell_T'X(X'X)^{-1}}{\ell_T'M\ell_T} \\ -\frac{(X'X)^{-1}X'\ell_T}{\ell_T'M\ell_T} & \left(X'X - \frac{X'\ell_T\ell_T'X}{T}\right)^{-1} \end{pmatrix} \begin{pmatrix} \ell_T \\ X' \end{pmatrix} \\
&= \frac{\ell_T\ell_T'}{\ell_T'M\ell_T} - \frac{X(X'X)^{-1}X'\ell_T\ell_T'}{\ell_T'M\ell_T} - \frac{\ell_T\ell_T'X(X'X)^{-1}X'}{\ell_T'M\ell_T} + X\left(X'X - \frac{X'\ell_T\ell_T'X}{T}\right)^{-1} X'
\end{aligned}
$$

Pluggin this into M_e yields

$$
M_e = I_T - \frac{\ell_T\ell_T'}{\ell_T'M\ell_T} + \frac{X(X'X)^{-1}X'\ell_T\ell_T'}{\ell_T'M\ell_T} + \frac{\ell_T\ell_T'X(X'X)^{-1}X'}{\ell_T'M\ell_T} - X\left(X'X - \frac{X'\ell_T\ell_T'X}{T}\right)^{-1} X'
$$

Therefore, the product of L and M_e becomes

$$
\begin{aligned}
LM_e &= \ell_N'(X'X)^{-1}X'M_e \\
&= \ell_N'(X'X)^{-1}X' - \frac{\ell_N'(X'X)^{-1}X'\ell_T\ell_T'}{\ell_T'M\ell_T} + \frac{\ell_N'(X'X)^{-1}X'\ell_T\ell_T'}{\ell_T'M\ell_T} \\
&\quad + \frac{\ell_N'(X'X)^{-1}X'\ell_T\ell_T'X(X'X)^{-1}X'}{\ell_T'M\ell_T} - \ell_N'\left(X'X - \frac{X'\ell_T\ell_T'X}{T}\right)^{-1} X' \\
&= \frac{1}{T}\ell_N'\hat{\Omega}_x^{-1}X' + \frac{\ell_N'\hat{\Omega}_x^{-1}\hat{\mu}_x\hat{\mu}_x'\hat{\Omega}_x^{-1}X'}{\ell_T'M\ell_T} - \frac{1}{T}\ell_N'\hat{\Sigma}_x^{-1}X' \\
&= \ell_N'\left[\frac{1}{T}\hat{\Omega}_x^{-1} + \frac{\hat{\Omega}_x^{-1}\hat{\mu}_x\hat{\mu}_x'\hat{\Omega}_x^{-1}}{\ell_T'M\ell_T} - \frac{1}{T}\hat{\Sigma}_x^{-1}\right] X'.
\end{aligned}
$$

In the above,

$$
\begin{aligned}
\ell_T'M\ell_T &= \ell_T'(I_T - X(X'X)^{-1}X')\ell_T \\
&= \ell_T'\ell_T - T^2\frac{\ell_T'X}{T}\left(\frac{X'X}{T}\cdot T\right)^{-1}\frac{X'\ell_T}{T} \\
&= T - T\hat{\mu}_x'\hat{\Omega}_x^{-1}\hat{\mu}_x \\
&= T\left(1 - \frac{\hat{a}_x}{1+\hat{a}_x}\right) \\
&= \frac{T}{1+\hat{a}_x},
\end{aligned}
\tag{59}
$$

which utilizes $\hat{\mu}_x'\hat{\Omega}_x^{-1}\hat{\mu}_x = \frac{\hat{a}_x}{1+\hat{a}_x}$ from Lemma A.3. Plugging this result back into LM_e yields

$$
LM_e = \frac{1}{T}\ell_N'\left[\hat{\Omega}_x^{-1} + (1+\hat{a}_x)\hat{\Omega}_x^{-1}\hat{\mu}_x\hat{\mu}_x'\hat{\Omega}_x^{-1} - \hat{\Sigma}_x^{-1}\right] X',
$$

where

$$\hat{\Omega}_x^{-1}\hat{\mu}_x = \left(\hat{\Sigma}_x^{-1} - \frac{\hat{\Sigma}_x^{-1}\hat{\mu}_x\hat{\mu}_x'\hat{\Sigma}_x^{-1}}{1+\hat{a}_x}\right)\hat{\mu}_x = \hat{\Sigma}_x^{-1}\hat{\mu}_x - \frac{\hat{\Sigma}_x^{-1}\hat{\mu}_x\hat{a}_x}{1+\hat{a}_x} = \frac{\hat{\Sigma}_x^{-1}\hat{\mu}_x}{1+\hat{a}_x}.$$

Finally,

$$\begin{aligned} LM_e &= \frac{1}{T}\ell_N'\left[\hat{\Omega}_x^{-1} + (1+\hat{a}_x)\frac{\hat{\Sigma}_x^{-1}\hat{\mu}_x\hat{\mu}_x'\hat{\Sigma}_x^{-1}}{(1+\hat{a}_x)^2} - \hat{\Sigma}_x^{-1}\right]X' \\ &= \frac{1}{T}\left[\hat{\Omega}_x^{-1} + \frac{\hat{\Sigma}_x^{-1}\hat{\mu}_x\hat{\mu}_x'\hat{\Sigma}_x^{-1}}{(1+\hat{a}_x)} - \hat{\Sigma}_x^{-1}\right]X,' \\ &= 0_T' \end{aligned}$$

where the last equality comes from the Sherman-Morrison formula. *q.e.d.*

Proof of Proposition 7-A: Below we derive the unbiased estimator of $\|e_c\|$, $\frac{\alpha_c^*}{\sigma_{e_c}}$ and $\alpha_c^*\sigma_{e_c}$.

(i) **The unbiased estimator of $\boldsymbol{\sigma_{e_c}}$**
Given that the unbiased estimator of $\sigma_{e_c}^2$ is $\hat{\sigma}_{e_c}^2 = \frac{\sum_{t=1}^T \hat{e}_{ct}^2}{T-N-1}$, which is χ^2 distributed:

$$\frac{(T-N-1)\hat{\sigma}_{e_c}^2}{\sigma_{e_c}^2} \sim \chi^2(T-N-1).$$

Then, its square-root has a χ distribution:

$$\sqrt{\frac{(T-N-1)\hat{\sigma}_{e_c}^2}{\sigma_{e_c}^2}} = \sqrt{T-N-1}\frac{\sqrt{\hat{\sigma}_{e_c}^2}}{\sigma_{e_c}} \sim \chi(T-N-1),$$

and its first moment is[31]

$$E\left[\sqrt{T-N-1}\frac{\sqrt{\hat{\sigma}_e^2}}{\sigma_{e_c}}\right] = \sqrt{2}\frac{\Gamma\left(\frac{T-N}{2}\right)}{\Gamma\left(\frac{T-N-1}{2}\right)}.$$

Therefore,

$$E\left[\frac{\sqrt{T-N-1}\,\Gamma\left(\frac{T-N-1}{2}\right)}{\sqrt{2}\,\Gamma\left(\frac{T-N}{2}\right)}\sqrt{\hat{\sigma}_{e_c}^2}\right] = \sigma_{e_c}.$$

Therefore, the unbiased estimator of σ_{e_c} is

$$s_{e_c} = \sqrt{\frac{T-N-1}{2}}\,\frac{\Gamma\left(\frac{T-N-1}{2}\right)}{\Gamma\left(\frac{T-N}{2}\right)}\sqrt{\hat{\sigma}_{e_c}^2}. \tag{60}$$

(ii) **The unbiased estimator of $\boldsymbol{\frac{\alpha_c^*}{\|e_c\|}}$**
From Proposition 6,

$$\frac{\hat{\alpha}_c^*}{\frac{\sigma_{e_c}}{\sqrt{T}}\sqrt{\frac{\hat{c}_x+\hat{a}_x\hat{c}_x-\hat{b}_x^2}{1+\hat{a}_x}}} \sim \mathcal{N}\left(\alpha_c^*, 1\right).$$

[31] See Gooch (2010).

From Equation(58), we know that

$$\frac{\hat{\alpha}_c^* - \alpha_c^*}{\sigma_{e_c}} = \ell_N'(X'X)^{-1}X'\left(\frac{\mathbf{e_c}}{\sigma_{e_c}}\right) \stackrel{let}{=} L\left(\frac{\mathbf{e_c}}{\sigma_{e_c}}\right). \tag{61}$$

On the other hand, the least squares residuals are $\hat{\boldsymbol{e}}_\mathbf{c} = M_e x_c = M_e \mathbf{e_c}$, where $X_e = (\ell_T \;\; X)$ and M_e is symmetric and imdemponent and its trace is $T - N - 1$. The well-known results from the standard least square regression yields

$$\frac{(T-N-1)\hat{\sigma}_{e_c}^2}{\sigma_{e_c}^2} = \left(\frac{\mathbf{e_c}}{\sigma_{e_c}}\right)'(I_T - X_e(X_e'X_e)^{-1}X_e')\left(\frac{\mathbf{e_c}}{\sigma_{e_c}}\right) \stackrel{let}{=} \left(\frac{\mathbf{e_c}}{\sigma_{e_c}}\right)' M_e \left(\frac{\mathbf{e_c}}{\sigma_{e_c}}\right). \tag{62}$$

(61) is a linear function of standard normal variates and (62) is their quadratic function. From Lemma A.4, $LM_e = 0$ and thus $L\left(\frac{e_c}{\sigma_{e_c}}\right)$ in (61) and $\left(\frac{\mathbf{e_c}}{\sigma_{e_c}}\right)' M_e \left(\frac{\mathbf{e_c}}{\sigma_{e_c}}\right)$ in (62) are statistically independent. As a result,

$$\frac{\hat{\alpha}_c^* - \alpha_c^*}{\frac{\sigma_{e_c}}{\sqrt{T}}\sqrt{\frac{\hat{c}_x + \hat{a}_x\hat{c}_x - \hat{b}_x^2}{1+\hat{a}_x}}} = \frac{\sqrt{T}}{\sqrt{\frac{\hat{c}_x + \hat{a}_x\hat{c}_x - \hat{b}_x^2}{1+\hat{a}_x}}} L\left(\frac{\mathbf{e_c}}{\sigma_{e_c}}\right)$$

is also independent of $\left(\frac{\mathbf{e_c}}{\sigma_{e_c}}\right)' M_e \left(\frac{\mathbf{e_c}}{\sigma_{e_c}}\right)$. Then the ratio of the non-zero mean normal variate with unit standard deviation to the χ distributed random variate becomes

$$\frac{\frac{\hat{\alpha}_c^*}{\frac{\sigma_{e_c}}{\sqrt{T}}\sqrt{\frac{\hat{c}_x + \hat{a}_x\hat{c}_x - \hat{b}_x^2}{1+\hat{a}_x}}}}{\sqrt{\frac{(T-N-1)\hat{\sigma}_e^2}{\sigma_{e_c}^2}}/\sqrt{T-N-1}} \sim t_{nc}\left(T-N-1, \frac{\alpha_c^*}{\frac{\sigma_{e_c}}{\sqrt{T}}\sqrt{\frac{\hat{c}_x + \hat{a}_x\hat{c}_x - \hat{b}_x^2}{1+\hat{a}_x}}}\right) \tag{63}$$

where $t_{nc}(K, \lambda)$ is a non-central t distribution with K degrees of freedom and non-centrality parameter λ.[32] Under the assumption of $T - N - 1 > 1$, we first simplify and take the expectation of the left-hand-side of (63):

$$E\left[\frac{\hat{\alpha}_c^*}{\sqrt{\frac{\hat{\sigma}_{e_c}^2}{T}}\sqrt{\frac{\hat{c}_x + \hat{a}_x\hat{c}_x - \hat{b}_x^2}{1+\hat{a}_x}}}\right] = \frac{\alpha_c^*}{\frac{\sigma_{e_c}}{\sqrt{T}}\sqrt{\frac{\hat{c}_x + \hat{a}_x\hat{c}_x - \hat{b}_x^2}{1+\hat{a}_x}}}\sqrt{\frac{T-N-1}{2}}\frac{\Gamma\left(\frac{T-N-2}{2}\right)}{\Gamma\left(\frac{T-N-1}{2}\right)}.$$

[32]The non-central t distribution is

$$t_{nc}(K, \lambda) = \frac{\mathcal{N}(\lambda, 1)}{\chi^2(K)/\sqrt{K}},$$

where $\mathcal{N}(\lambda, 1)$ is a normal variate with mean μ and unit variance whereas $\chi(K)$ is a χ distribution with K degrees of freedom. The ratio of the normal variate with non-zero mean with the χ variate divided by its square-roof of degrees of freedom is a non-central t distribution. Its square, $t_{nc}(K, \lambda)^2$ is $F(1, K, \lambda^2)$, a non-central F distribution with 1 numerator degree of freedom, K denominator degrees of freedom and non-centrality parameter λ^2. When $K > 1$, the first moment of a non-central t variate exists and it is

$$E[t_{nc}(K, \lambda)] = \lambda\sqrt{\frac{K}{2}}\frac{\Gamma(\frac{K-1}{2})}{\Gamma(\frac{K}{2})}.$$

See Cousineau and Laurencelle (2011) for details.

This can be rewritten as:

$$E\left[\sqrt{\frac{2}{T-N-1}}\frac{\Gamma\left(\frac{T-N-1}{2}\right)}{\Gamma\left(\frac{T-N-2}{2}\right)}\frac{\hat{\alpha}_c^*}{\sqrt{\hat{\sigma}_{e_c}^2}}\right]=\frac{\alpha_c^*}{\sigma_{e_c}}.$$

Therefore, the unbiased estimator of $\frac{\alpha_c^*}{\|e_c\|}$ is

$$\widehat{\left(\frac{\alpha_c^*}{\sigma_{e_c}}\right)}=\sqrt{\frac{2}{T-N-1}}\frac{\Gamma\left(\frac{T-N-1}{2}\right)}{\Gamma\left(\frac{T-N-2}{2}\right)}\frac{\hat{\alpha}_c^*}{\sqrt{\hat{\sigma}_{e_c}^2}}.$$

One can rewrite the above result as an expression of s_{e_c} using the relation between $\hat{\sigma}_{e_c}^2$ and s_{e_c}.

(iii) **The unbiased estimator of $\boldsymbol{\alpha_c^*\|e_c\|}$**
In (ii), we show that α_c^* is independent of $\hat{\sigma}_{e_c}^2$. Since $s_{e_c}^2$ is a linear function of $\hat{\sigma}_{e_c}^2$, α_c^* is also independent of $s_{e_c}^2$. If two random variables are independent, the functions of each random variables are also independent. As a result,

$$E\left(\hat{\alpha}_c^*\sqrt{s_{e_c}^2}\right)=E\left(\hat{\alpha}_c^*\right)E\left(\sqrt{s_{e_c}^2}\right)=E(\hat{\alpha}_c^*)E(s_{e_c})=\alpha_c^*\sigma_{e_c}.$$

Therefore the unbiased estimator of $\alpha_c^*\sigma_{e_c}$ is

$$\widehat{\alpha_c^*\sigma_{e_c}}=\hat{\alpha}_c^*s_{e_c}. \qquad q.e.d.$$

Proof of Proposition 7-B: Below we derive the pseudo-unbiased estimator of $\|\omega_c\|$, $\frac{\alpha_c^*}{\|\omega_c\|}$ and $\alpha_c^*\|\omega_c\|$. Before doing so, we first derive the unbiased estimator of $\|\omega_c\|^2=E(\omega_c^2)$ in two different ways. From Lemma A.2, $\|\omega_c\|^2=\sigma_{e_c}^2+\frac{\beta_{c0}^2}{1+a_x}$. Recall that the unbiased estimator of $\sigma_{e_c}^2$ is $\hat{\sigma}_{e_c}^2=\frac{\hat{\mathbf{e}}_{\mathbf{c}}'\hat{\mathbf{e}}_{\mathbf{c}}}{T-N-1}$. The least sqaures residuals in the absence of the intercept are

$$\begin{aligned}\hat{\boldsymbol{\omega}}_{\mathbf{c}} &= X_c-X\hat{b}_c\\ &= X_c-X(X'X)^{-1}X'X_c\\ &= (I_T-X(X'X)^{-1}X')X_c\\ &= MX_c\\ &= M(\beta_{c0}\ell_T+X\beta_c+\mathbf{e_c})\\ &= \beta_{c0}M\ell_T+M\mathbf{e_c},\end{aligned} \tag{64}$$

since $MX=0_{T\times N}$. Then $\hat{\boldsymbol{\omega}}_{\mathbf{c}}'\ell_T=\beta_{c0}\ell_T'M\ell_T+\mathbf{e_c}'M\ell_T$, and as a result,

$$\frac{\hat{\boldsymbol{\omega}}_{\mathbf{c}}'\ell_T}{\ell_T'M\ell_T}=\beta_{c0}+\frac{\ell_T'M\mathbf{e_c}}{\ell_T'M\ell_T}.$$

Further, $\mathrm{var}\left(\frac{\ell_T'M\mathbf{e_c}}{\ell_T'M\ell_T}\right)=\sigma_{e_c}^2\frac{\ell_T'MI_TM\ell_T}{(\ell_T'M\ell_T)^2}=\frac{\sigma_{e_c}^2}{\ell_T'M\ell_T}$ because M is idempotent. As shown in (59), $\ell_T'M\ell_T=\frac{T}{1+\hat{a}_x}$. In sum,

$$\hat{\beta}_{c0}=\frac{\hat{\boldsymbol{\omega}}_{\mathbf{c}}'\ell_T}{\ell_T'M\ell_T}\sim\mathcal{N}\left(\beta_{c0},\frac{1+\hat{a}_x}{T}\sigma_{e_c}^2\right).$$

As a result, $E(\hat{\beta}_{c0}^2) = \beta_{c0}^2 + \frac{1+\hat{a}_x}{T}\sigma_{e_c}^2$, and thus $\hat{\beta}_{c0}^2$ is a biased estimator of β_{c0}^2 unless β_{c0} is zero. Because we are interested not only in testing $\beta_{c0} = 0$, but also in measuring the degree of domestication. As such, we need to take into account the bias driven by non-zero β_{c0}. To correct for such a bais, we subtract it by the unbiased estimator of $\frac{1+hata_x}{T}\sigma_{e_c}^2$ such that

$$E\left[\hat{\beta}_{c0}^2 - \frac{1+\hat{a}_x}{T}\hat{\sigma}_{e_c}^2\right] = \beta_{c0}^2.$$

Finally we make a sum of the two unbiased estimators,

$$\hat{\sigma}_{\omega_c}^2 = \widehat{\|\omega_c\|^2} = \hat{\sigma}_e^2 + \frac{\hat{\beta}_{c0}^2 - \frac{1+\hat{a}_x}{T}\hat{\sigma}_e^2}{1+\hat{a}_x} = \frac{\hat{\beta}_{c0}^2}{1+\hat{a}_x} + \frac{T-1}{T}\hat{\sigma}_{e_c}^2, \tag{65}$$

and $E(\hat{\sigma}_{\omega_c}^2) = \sigma_{e_c}^2 + \frac{\beta_{c0}^2}{1+\hat{a}_x}$, which is unbiased. For a use later, we derive this unbiased estimator by an alternative method. We begin with $\hat{\boldsymbol{\omega}}_{\boldsymbol{c}} = \beta_{c0}M\ell_T + M\mathbf{e_c} \sim N_T(\beta_{c0}M\ell_T, \sigma_{e_c}^2 M)$ in (64). Given that the trace of M is $T-N$, its quadratic form:

$$\hat{\boldsymbol{\omega}}_{\boldsymbol{c}}'(\sigma_{e_c}^2 M)^{-1}\hat{\boldsymbol{\omega}}_{\boldsymbol{c}} \sim \chi_{nc}^2\left(T-N, \frac{\beta_{c0}^2\ell_T' M\ell_T}{\sigma_{e_c}^2}\right), \tag{66}$$

which is a noncentral χ^2 distribution with $T-N$ degrees of freedom and noncentrality parameter $\beta_{c0}^2\ell_T' M\ell_T/\sigma_{e_c}^2$. In addition, it is easy to show that $\hat{\boldsymbol{\omega}}_{\boldsymbol{c}}' M^{-1}\hat{\boldsymbol{\omega}}_{\boldsymbol{c}} = \hat{\boldsymbol{\omega}}_{\boldsymbol{c}}'\hat{\boldsymbol{\omega}}_{\boldsymbol{c}}$. The expected value of the noncentral χ^2 is the sum of the degrees of freedom and the noncentrality parameters, which yields

$$E\left(\frac{1}{\sigma_{e_c}^2}\hat{\boldsymbol{\omega}}_{\boldsymbol{c}}'\hat{\boldsymbol{\omega}}_{\boldsymbol{c}}\right) = T-N+\frac{\beta_{c0}^2\ell_T' M\ell_T}{\sigma_{e_c}^2}.$$

Since $\ell_T' M\ell_t = T/(1+\hat{a}_x)$, $E(\hat{\boldsymbol{\omega}}_{\boldsymbol{c}}'\hat{\boldsymbol{\omega}}_{\boldsymbol{c}}) = \beta_{c0}^2\frac{T}{1+a_x} + (T-N)\sigma_{e_c}^2$. Therefore, $\hat{\boldsymbol{\omega}}_{\boldsymbol{c}}'\hat{\boldsymbol{\omega}}_{\boldsymbol{c}}/T$ is a consistent estimator of $E(\omega^2)$, but it is biased. To correct for such a bias, we add $\frac{N}{T}\hat{\sigma}_{e_c}^2$. Then,

$$E\left(\frac{\hat{\boldsymbol{\omega}}_{\boldsymbol{c}}'\hat{\boldsymbol{\omega}}_{\boldsymbol{c}}}{T} + \frac{N}{T}\hat{\sigma_{e_c}}^2\right) = \frac{\beta_{c0}^2}{1+a_x} + \frac{T-N}{T}\sigma_{e_c}^2 + \frac{N}{T}\sigma_{e_c}^2 = \frac{\beta_{c0}^2}{1+a_x} + \sigma_{e_c}^2.$$

Therefore, the unbiased estimator is

$$\hat{\sigma}_{\omega_c}^2 = \widehat{\|\omega_c\|^2} = \frac{\hat{\boldsymbol{\omega}}_{\boldsymbol{c}}'\hat{\boldsymbol{\omega}}_{\boldsymbol{c}}}{T} + \frac{N}{T}\hat{\sigma}_{e_c}^2. \tag{67}$$

Given the definition of $\hat{\boldsymbol{\omega}}_{\boldsymbol{c}}$, it is, albeit tedious, straightfoward to prove the equality of (65) and (67).

(i) **The pseudo-unbiased estimator of $\|\boldsymbol{\omega_c}\| = \sqrt{\boldsymbol{E}(\boldsymbol{\omega^2})}$**
Given the unbiased estimator $\|\omega_c\|^2$,

$$\begin{aligned}
\frac{T\widehat{\|\omega_c\|^2}}{\sigma_{e_c}^2} &= \frac{\hat{\boldsymbol{\omega}}_{\boldsymbol{c}}'\hat{\boldsymbol{\omega}}_{\boldsymbol{c}}}{\sigma_{e_c}^2} + N\frac{\hat{\sigma_{e_c}}^2}{\sigma_{e_c}^2} \\
&= \frac{\mathbf{e_c}'}{\sigma_{e_c}}\left(M + \frac{N}{T-N-1}M_e\right)\frac{\mathbf{e_c}}{\sigma_{e_c}} + 2\frac{\beta_{c0}\ell_T' M}{\sigma_{e_c}}\frac{\mathbf{e_c}}{\sigma_{e_c}} + \frac{\beta_{c0}^2\ell_T' M\ell_T}{\sigma_{e_c}^2},
\end{aligned}$$

which is a generalized chi-squared variable. As far as we know, the expected value of its square root is yet to be found. Even in the case that we can get it, it does not guarantee the existence of the unbiased estimator of $\|\omega_c\|$. As such, we may rely upon a simulation to correct for a Jensen's inequality stemming from the square root, which is a nonlinear function. As an alternative, here we take a different approach to approximate it. Taking a square root on (66) yields

$$\sqrt{\frac{\hat{\boldsymbol{\omega}}_{\boldsymbol{c}}{}'\hat{\boldsymbol{\omega}}_{\boldsymbol{c}}}{\sigma_{e_c}^2}} \sim \chi_{nc}\left(T-N, \sqrt{\frac{\beta_{c0}^2\ell_T' M\ell_T}{\sigma_{e_c}^2}}\right), \tag{68}$$

which is a noncentral chi distribution with $T-N$ degrees of freedom and centrality parameter, $\sqrt{\frac{\beta_{c0}^2\ell_T' M\ell_T}{\sigma_{e_c}^2}}$. We construct an estimator, the degrees of freedom of which are consistent with the χ_{nc} distribution:

$$\hat{\hat{\sigma}}_{\omega_c}^2 = \frac{\hat{\boldsymbol{\omega}}_{\boldsymbol{c}}{}'\hat{\boldsymbol{\omega}}_{\boldsymbol{c}}}{T-N}. \tag{69}$$

Then, the first moment of the χ_{nc} distribution is

$$E\left(\sqrt{\frac{\hat{\boldsymbol{\omega}}_{\boldsymbol{c}}{}'\hat{\boldsymbol{\omega}}_{\boldsymbol{c}}}{\sigma_{e_c}^2}}\right) = E\left(\sqrt{\frac{(T-N)\hat{\hat{\sigma}}_{\omega_c}^2}{\sigma_{e_c}^2}}\right) = \sqrt{\frac{\pi}{2}}\; L_{1/2}^{\frac{T-N}{2}-1}\left(-\frac{\beta_{c0}^2\ell_T' M\ell_T}{2\sigma_{e_c}^2}\right), \tag{70}$$

where $L_{\theta_l}^{\theta_u}(\cdot)$ is a generalized Laguerre fuction.[33] Let $\theta = \frac{\beta_{c0}^2\ell_T' M\ell_T}{\sigma_{e_c}^2}$, the centrality parameter of the χ_{nc}^2 distribution. Given the well-known relationship between the generalized Laguerre function and the confluent hypergeometric function of the first kind, $L_p^q(y) = \frac{(q+1)_p}{p!}\,{}_1F_1(-p;q+1;y)$ (where $(a)_b$ is the Pochhammer symbol), we can rewrite (70):

$$\begin{aligned}
E\left(\sqrt{\frac{(T-N)\hat{\hat{\sigma}}_{\omega_c}^2}{\sigma_{e_c}^2}}\right) &= \sqrt{\frac{\pi}{2}}\; L_{1/2}^{\frac{T-N}{2}-1}\left(-\frac{\theta}{2}\right) \\
&= \sqrt{\frac{\pi}{2}}\; \frac{\left(\frac{T-N}{2}\right)_{1/2}}{\frac{1}{2}!}\; {}_1F_1\left(-\frac{1}{2};\frac{T-N}{2};-\frac{\theta}{2}\right) \\
&= \sqrt{\frac{\pi}{2}}\; \frac{\Gamma\left(\frac{T-N+1}{2}\right)/\Gamma\left(\frac{T-N}{2}\right)}{\frac{\sqrt{\pi}}{2}}\; {}_1F_1\left(-\frac{1}{2};\frac{T-N}{2};-\frac{\theta}{2}\right) \\
&= \sqrt{2}\frac{\Gamma\left(\frac{T-K+1}{2}\right)}{\Gamma\left(\frac{T-N}{2}\right)}{}_1F_1\left(-\frac{1}{2};\frac{T-N}{2};-\frac{\theta}{2}\right),
\end{aligned} \tag{71}$$

where we use $(a)_b = \Gamma(a+b)/\Gamma(a)$, and $\frac{1}{2}! = \Gamma(\frac{3}{2}) = \frac{\sqrt{\pi}}{2}$. The confluent hypergeometric function, ${}_1F_1(\cdot;\cdot;\cdot)$ has a hypergeometric series given by

$$\begin{aligned}
{}_1F_1\left(-\frac{1}{2};\frac{T-N}{2};-\frac{\theta}{2}\right) &= \sum_{j=0}^{\infty}\frac{\left(-\frac{1}{2}\right)_j}{\left(\frac{T-N}{2}\right)_j}\frac{\left(-\frac{\theta}{2}\right)^j}{j!} \\
&= 1+\frac{1}{T-N}\left(\frac{\theta}{2}\right)-\frac{1}{(T-N)(T-N+2)}\left(\frac{\theta}{2}\right)^2+\cdots \\
&\approx 1+\frac{\theta}{2(T-N)},
\end{aligned}$$

[33] See Park (1961) for the moments of a noncentral χ distribution.

The final equation is based on the first order approximation, which is quite accurate because

$$\theta = \frac{\beta_{c0}^2 \ell_N' M \ell_N}{\sigma_{e_c}^2} = \frac{\beta_{c0}^2 T}{(1+\hat{a}_x)\sigma_{e_c}^2} = \frac{\beta_{c0}^2 T}{(1+\hat{\mu}_x'\hat{\Sigma}_x^{-1}\hat{\mu}_x)\sigma_{e_c}^2},$$

which is quite small since $\hat{a}_x = \hat{\mu}_x'\hat{\Sigma}_x^{-1}\hat{\mu}_x$ is a fairly large value. Then, (71) can be rewritten as:

$$E\left(\sqrt{\hat{\hat{\sigma}}_{\omega_c}^2}\right) \approx \frac{\sqrt{2}}{\sqrt{T-N}} \frac{\Gamma\left(\frac{T-K+1}{2}\right)}{\Gamma\left(\frac{T-N}{2}\right)} \sigma_{e_c} \left(1 + \frac{\theta}{2(T-N)}\right), \tag{72}$$

where

$$\sigma_{e_c}\left(1 + \frac{\theta}{2(T-N)}\right) = \sigma_{e_c} + \frac{T}{2(T-N)} \frac{\beta_{c0}^2}{(1+\hat{a}_x)\sigma_{e_c}}.$$

The first-order Taylor approximation of $\sqrt{a+b}$ at $b=0$ is $\sqrt{a} + \frac{1}{2\sqrt{a}}b$. Therefore,

$$\sigma_{e_c} + \frac{T}{2(T-N)} \frac{\beta_{c0}^2}{(1+\hat{a}_x)\sigma_{e_c}} \approx \sqrt{\sigma_{e_c}^2 + \frac{T}{T-N}\frac{\beta_{c0}^2}{1+\hat{a}_x}}. \tag{73}$$

Consequently,

$$E\left(\sqrt{\hat{\hat{\sigma}}_{\omega_c}^2}\right) \approx \frac{\sqrt{2}}{\sqrt{T-N}} \frac{\Gamma\left(\frac{T-K+1}{2}\right)}{\Gamma\left(\frac{T-N}{2}\right)} \sqrt{\sigma_{e_c}^2 + \frac{T}{T-N}\frac{\beta_{c0}^2}{1+a_x}}$$

Therefore, the unbiased estimator of $\sqrt{\sigma_{e_c}^2 + \frac{T}{T-N}\frac{\beta_{c0}^2}{1+a_x}}$ is

$$\hat{\hat{s}}_{\omega_c} = \frac{\sqrt{T-N}}{\sqrt{2}} \frac{\Gamma\left(\frac{T-N}{2}\right)}{\Gamma\left(\frac{T-N+1}{2}\right)} \sqrt{\hat{\hat{\sigma}}_{\omega_c}^2} = \frac{\sqrt{T-N}}{\sqrt{2}} \frac{\Gamma\left(\frac{T-N}{2}\right)}{\Gamma\left(\frac{T-N+1}{2}\right)} \sqrt{\frac{\hat{\boldsymbol{\omega}}_{\boldsymbol{c}}'\hat{\boldsymbol{\omega}}_{\boldsymbol{c}}}{T-N}}. \tag{74}$$

The problem of $\hat{\hat{s}}_{\omega_c}$ is its expectation is not precisely equal to $\|\omega_c\|$:

$$E(\hat{\hat{s}}_{\omega_c}) \approx \sqrt{\sigma_{e_c}^2 + \frac{T}{T-N}\frac{\beta_{c0}^2}{1+\hat{a}_x}} \neq \sqrt{\sigma_{e_c}^2 + \frac{\beta_{c0}^2}{1+\hat{a}_x}} = \sqrt{E(\omega^2)}.$$

In (74), $\sqrt{\frac{T-N}{2}}\Gamma(\frac{T-N}{2})/\Gamma(\frac{T-N+1}{2})$ is the bias adjustment factor, which is precisely identical to the bias adjustment in s_{e_c}, $\sqrt{\frac{T-N-1}{2}}\Gamma(\frac{T-N-1}{2})/\Gamma(\frac{T-N}{2})$ in Equation (60) after an adjustment of the degrees of freedom. As such, we can conjecture that $\sqrt{\frac{T-N}{2}}\Gamma(\frac{T-N}{2})/\Gamma(\frac{T-N-1}{2})$ is a bias adjustment factor which is required for adjusting the Jensen's inequality associated with square-root transformation. Following this intuition, we consider the following estimator of $\|\omega_c\|$:

$$s_{\omega_c} = \frac{\sqrt{T-N}}{\sqrt{2}} \frac{\Gamma\left(\frac{T-N}{2}\right)}{\Gamma\left(\frac{T-N+1}{2}\right)} \sqrt{\hat{\sigma}_{\omega_c}^2} = \frac{\sqrt{T-N}}{\sqrt{2}} \frac{\Gamma\left(\frac{T-N}{2}\right)}{\Gamma\left(\frac{T-N+1}{2}\right)} \sqrt{\frac{\hat{\boldsymbol{\omega}}_{\boldsymbol{c}}'\hat{\boldsymbol{\omega}}_{\boldsymbol{c}}}{T} + \frac{N}{T}\hat{\sigma}_e^2},$$

i.e., we use $\hat{\sigma}_{\omega_c}^2$, the unbiased estimator of $\|\omega_c\|^2$ instead of $\hat{\hat{\sigma}}_{\omega_c}^2$. The simulation analysis shows that s_{ω_c}, a pseudo-unbiased estimator, is, in fact, almost unbiased: $E(s_{\omega_c}) = \sqrt{\sigma_{e_c}^2 + \frac{\beta_{c0}^2}{1+a_x}}$. Table A.1. summarizes simulation results on the unbiasedness of s_w.

(ii) **The pseudo-unbiased estimator of $\frac{\boldsymbol{\alpha_x^*}}{\|\boldsymbol{\omega_c}\|}$:**
Given the distribution of $\hat{\alpha}_c^*$, (34) in Proposition 6 and that of $\hat{\hat{\sigma}}^2_{\omega_c}$ in (68) and (69) in (i) of this proposition, we consider the following statistic:

$$t''_{nc}(\hat{\alpha}_c^*, \hat{\hat{\sigma}}^2_{\omega_c}) = \frac{\frac{\hat{\alpha}_c^*}{\frac{\sigma_{e_c}}{\sqrt{T}}\sqrt{\frac{\hat{c}_x+\hat{a}_x\hat{c}_x-\hat{b}_x^2}{1+\hat{a}_x}}}}{\sqrt{\frac{(T-N)\hat{\hat{\sigma}}^2_{\omega_c}}{\sigma^2_{e_c}}}/\sqrt{T-N}} = \frac{\hat{\alpha}_c^*}{\frac{\sqrt{\hat{\hat{\sigma}}^2_{\omega_c}}}{\sqrt{T}}\sqrt{\frac{\hat{c}_x+\hat{a}_x\hat{c}_x-\hat{b}_x^2}{1+\hat{a}_x}}} \sim t''_{nc}(\nu, \delta, \theta), \tag{75}$$

where

$$\nu = T - N, \quad \delta = \frac{\alpha_c^*}{\frac{\sigma_{e_c}}{\sqrt{T}}\sqrt{\frac{\hat{c}_x+\hat{a}_x\hat{c}_x-\hat{b}_x^2}{1+\hat{a}_x}}}, \quad \theta = \frac{\beta_{c0}^2 T}{(1+\hat{a}_x)\sigma^2_{e_c}}.$$

$t''_{nc}(\nu, \delta, \theta)$ denotes a doubly noncentral t distribution, which is isomorphic to the standard central t distribution or the noncentral t distribution examined in Proposition 7-A. It is a ratio of two random variables where its numerator is a normal variate with mean δ and unit variance and its denominator is a square-root of an independent noncentral chi-square variate with ν degrees of freedom and noncentrality parameter θ. Using the idemponency of M in $\hat{\omega}_{\mathbf{c}}$, we can prove that the numerator is independent of the nominator in (75) in a similar way shown in Lemma A.4. Its first moment is $E(t''_{nc}(\nu, \delta, \theta)) = \delta\sqrt{\frac{\nu}{2}}\frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}{}_1F_1\left(\frac{1}{2}, \frac{\nu}{2}, -\frac{\theta}{2}\right)$ when $\nu > 1$.[34] Therefore, the expected value of t-static in (75) is

$$E\big[t''(\alpha_c^*, \hat{\hat{\sigma}}^2_{\omega_c})\big] = \frac{\alpha_c^*}{\frac{\sigma_{e_c}}{\sqrt{T}}\sqrt{\frac{\hat{c}_x+\hat{a}_x\hat{c}_x-\hat{b}_x^2}{1+\hat{a}_x}}}\sqrt{\frac{T-N}{2}}\frac{\Gamma\left(\frac{T-N-1}{2}\right)}{\Gamma\left(\frac{T-N}{2}\right)}{}_1F_1\left(\frac{1}{2}, \frac{T-N}{2}, -\frac{\theta}{2}\right),$$

which yields

$$E\left[\frac{\hat{\alpha}_c^*}{\sqrt{\hat{\hat{\sigma}}^2_{\omega_c}}}\right] = \sqrt{\frac{T-N}{2}}\frac{\Gamma\left(\frac{T-N-1}{2}\right)}{\Gamma\left(\frac{T-N}{2}\right)}\left[\frac{\alpha_c^*}{\sigma_{e_c}}\,{}_1F_1\left(\frac{1}{2}, \frac{T-N}{2}, -\frac{\theta}{2}\right)\right].$$

From the first-order approximation of the confluent hypergeomtric function, we get

$$\frac{\alpha_c^*}{\sigma_{e_c}}\,{}_1F_1\left(\frac{1}{2}, \frac{T-N}{2}, -\frac{\theta}{2}\right) \approx \frac{\alpha_c^*}{\sigma_{e_c}}\left(1 - \frac{1}{T-N}\cdot\frac{\theta}{2}\right) = \frac{\alpha_c^*}{\sigma_{e_c}/\left(1 - \frac{1}{T-N}\cdot\frac{\theta}{2}\right)}.$$

The first-order Taylor expansion gives $1/(1-a) \approx 1+a$, and thus we get $\sigma_{e_c}/\left(1 - \frac{1}{T-N}\cdot\frac{\theta}{2}\right) \approx \sigma_{e_c}\left(1 + \frac{1}{T-N}\frac{\theta}{2}\right) \approx \sqrt{\sigma^2_{e_c} + \frac{T}{T-N}\frac{\beta^2_{c0}}{1+a_x}}$. The last approximation is a deja vu of (73). Consequently,

$$E\left[\frac{\hat{\alpha}_c^*}{\sqrt{\hat{\hat{\sigma}}^2_{\omega_c}}}\right] = \sqrt{\frac{T-N}{2}}\frac{\Gamma\left(\frac{T-N-1}{2}\right)}{\Gamma\left(\frac{T-N}{2}\right)}\left[\frac{\alpha_c^*}{\sqrt{\sigma^2_{e_c} + \frac{T}{T-N}\frac{\beta^2_{c0}}{1+a_x}}}\right].$$

[34] See Krishnan (1967) for details.

Again,

$$\sqrt{\frac{2}{T-N}}\frac{\Gamma\left(\frac{T-N}{2}\right)}{\Gamma\left(\frac{T-N-1}{2}\right)}E\left(\frac{\hat{\alpha}_c^*}{\sqrt{\hat{\hat{\sigma}}^2_{\omega_c}}}\right)=\frac{\alpha_c^*}{\sqrt{\sigma^2_{e_c}+\frac{T}{T-N}\frac{\beta_{c0}^2}{1+a_x}}},$$

which is different from $\alpha_c^*/\sqrt{\sigma^2_{e_c}+\frac{\beta_{c0}^2}{1+a_x}}$ thereby being biased. As in (i) above, to correct for such a bias, we consider the following the following pseudo-unbiased estimator:

$$\widehat{\left(\frac{\alpha_c^*}{\|\omega_c\|}\right)}=\sqrt{\frac{2}{T-N}}\frac{\Gamma\left(\frac{T-N}{2}\right)}{\Gamma\left(\frac{T-N-1}{2}\right)}\left(\frac{\hat{\alpha}_c^*}{\sqrt{\hat{\sigma}^2_{\omega_c}}}\right),$$

that is, we use $\hat{\sigma}^2_{\omega_c}$, the unbiased estimator of $\|\omega_c\|^2$ instead of $\hat{\hat{\sigma}}^2_{\omega_c}$. A simulation result tabulated in Table A.1. shows that, in fact, this estimator is almost unbiased.

(iii) **The unbiased estimator of $\boldsymbol{\alpha_x^*\|\omega_c\|}$**:
Since $\hat{\alpha}_c^*$ and s_{w_c} are independent, we simply make a product of the two estimators:

$$\widehat{\alpha_c^*\|\omega_c\|}=\hat{\alpha}_c^* s_{\omega_c}.$$

This completes the proof. *q.e.d.*

B MCMC Algorithm

B.1 State Space Form

The DCF model we estimate a state-space model with a Markov-regime switching parameters. In the model the Markov process M_t evolves according to a transition probability matrix,

$$P = \begin{pmatrix} p_{00} & p_{01} = 1 - p_{00} \\ p_{10} = 0 & p_{11} = 1 \end{pmatrix},$$

where the initial state is 0. Given the state at time t, the measurement equation in Equation (1) can be expressed in a matrix form,

$$Y_t = H_{M_t} \times F_t + E_t,$$

where $Y_t = (S_t', A_t, D_t, B_t)'$ is the vector of the observations, $F_t = (G_t, L_t^S, L_t^A, L_t^D)'$ is the vector of the common factors, $E_t = (E_t^S, E_t^A, E_t^D, E_t^B)'$ is the vector of the measurement errors, and

$$H_{M_t} = \begin{pmatrix} \gamma_{S,M_t} & \lambda_{S,M_t} & 0_{3\times 1} & 0_{3\times 1} \\ \gamma_{A,M_t} & 0 & \delta_{A,M_t} & 0 \\ \gamma_{D,M_t} & 0 & 0 & \kappa_{D,M_t} \\ \gamma_{B,M_t} & \lambda_{B,M_t} & \delta_{B,M_t} & \kappa_{B,M_t} \end{pmatrix}$$

is the matrix of the factor loadings. The transition equation is given in Equation (2). We follow the approach of Kim and Kang (2019) and compute the likelihood using the Markov state process and the measurement and transition equations. As all priors are conjugate, we compute and compare the marginal likelihoods of the models with different number of changepoints using the method of Chib (1995).

The quantities to be sampled are the model parameters

$$\begin{aligned} \theta \quad = \quad & (P, \{\gamma_{S,M_t=j}, \gamma_{A,M_t=j}, \gamma_{D,M_t=j}, \gamma_{B,M_t=j}, \lambda_{S,M_t=j}, \lambda_{B,M_t=j}, \delta_{A,M_t=j}, \delta_{B,M_t=j}, \kappa_{D,M_t=j}, \kappa_{B,M_t=j}\}_{j=1}^{1}, \\ & \sigma_G^2,\ \sigma_S^2,\ \sigma_A^2,\ \sigma_D^2,\ \Sigma_{M_t},\ \sigma_{B,M_t}^2), \end{aligned}$$

common factors ($\mathbf{F} = \{F_t\}_{t=1}^T$), and regime process ($\mathbf{M} = \{M_t\}_{t=1}^T$). To identify the common factors and states, we impose the following restrictions: for all M_t, δ_{A,M_t}, κ_{D,M_t} and the first elements of γ_{S,M_t} and λ_{S,M_t} are equal to one.

The first step of the posterior sampling through a Markov chain Monte Carlo (MCMC) simulation is to initialize $\mathbf{M}$ and θ. The first half of the sample period is set to be state 0, and the second half is state 1. θ is initialized at their prior means. Then, we sequentially sample $\mathbf{F}$, θ and $\mathbf{M}$ from their full conditional distributions in each MCMC iteration. The following describes the details of block-wise sampling.

B.1.1 Sampling F

We denote the conditional variance-covariance of Y_t by $\Lambda_{M_t} = \begin{bmatrix} \Sigma_{M_t} & 0_{5\times1} \\ 0_{1\times5} & \sigma^2_{B,M_t} \end{bmatrix}$. Then, using the joint conditional normal distribution of (F_t, Y_t),

$$\begin{pmatrix} F_t \\ Y_t \end{pmatrix} |\mathscr{Y}_{t-1}, \theta, \mathbf{M} \sim \mathscr{N}\left(0, \begin{pmatrix} \Omega_t & \Omega_t H'_{M_t} \\ H_{M_t}\Omega_t & H_{M_t}\Omega_t H'_{M_t} + \Lambda_{M_t} \end{pmatrix}\right),$$

and the assumption that the factors are serially uncorrelated given the parameters and states, we can derive the full conditional distribution of F_t as

$$F_t|\mathscr{Y}_t, \theta, \mathbf{M} \sim \mathscr{N}\left(\Omega_t H'_{M_t}\left(H_{M_t}\Omega_t H'_{M_t} + \Lambda_{M_t}\right)^{-1} Y_t,\ \Omega_t - \Omega_t H'_{M_t}\left(H_{M_t}\Omega_t H'_{M_t} + \Lambda_{M_t}\right)^{-1} H_{M_t}\Omega_t\right).$$

Then, F_t is drawn from this distribution.

B.1.2 Sampling $(\sigma^2_G, \sigma^2_S, \sigma^2_A, \sigma^2_D)$

As the common factors are serially uncorrelated and mutually independent, we can simply sample each factor variance from inverse gamma distributions as the following:

$$\begin{aligned}
\sigma^2_G|\mathbf{F} &\sim IG\left(\frac{\alpha_0 + T}{2}, \frac{\delta_0 + \sum_{t=1}^T G_t^2}{2}\right), \\
\sigma^2_S|\mathbf{F} &\sim IG\left(\frac{\alpha_0 + T}{2}, \frac{\delta_0 + \sum_{t=1}^T L_t^{S2}}{2}\right), \\
\sigma^2_A|\mathbf{F} &\sim IG\left(\frac{\alpha_0 + T}{2}, \frac{\delta_0 + \sum_{t=1}^T L_t^{A2}}{2}\right), \text{ and} \\
\sigma^2_D|\mathbf{F} &\sim IG\left(\frac{\alpha_0 + T}{2}, \frac{\delta_0 + \sum_{t=1}^T L_t^{D2}}{2}\right).
\end{aligned}$$

B.1.3 Sampling $(\gamma_{S,M_t}, \lambda_{S,M_t}, \gamma_{A,M_t}, \delta_{A,M_t}, \gamma_{D,M_t}, \kappa_{D,M_t}, \Sigma_{M_t})$

Given the common factors and states, the parameters in the measurement equations of $Q_t = (S'_t, A'_t, D'_t)'$, given by

$$Q_t|\theta, \mathbf{F}, M_t \sim \mathscr{N}\left(x_t \times \beta_{M_t}, \Sigma_{M_t}\right),$$

can be updated in a seeming unrelated regression model framework, where

$$x_t = \begin{pmatrix} G_t \times I_3 & L_t^S \times I_3 & 0_{3\times1} & 0_{3\times1} & 0_{3\times1} & 0_{3\times1} \\ 0_{1\times3} & 0_{1\times3} & G_t & L_t^A & 0 & 0 \\ 0_{1\times3} & 0_{1\times3} & 0 & 0 & G_t & L_t^D \end{pmatrix}$$

is a matrix of independent variables and

$$\beta_{M_t} = \begin{pmatrix} \gamma_{S,M_t} \\ \lambda_{S,M_t} \\ \gamma_{A,M_t} \\ \delta_{A,M_t} \\ \gamma_{D,M_t} \\ \kappa_{D,M_t} \end{pmatrix}$$

is the vector of the factor loadings. Imposing the factor and state identification restrictions, β_{M_t} is sampled from its full conditional distribution,

$$\mathcal{N}\left(\left(B_0^{-1} + \sum_{t=1}^{T} x_t \Sigma_{M_t}^{-1} x_t'\right)^{-1} \left(\sum_{t=1}^{T} x_t \Sigma_{M_t}^{-1} Q_t\right), \left(B_0^{-1} + \sum_{t=1}^{T} x_t x_t'\right)^{-1}\right),$$

where $B_0 = V_{\beta_c} \times I_{\dim(\beta_{M_t})}$ is the the prior variance of β_{M_t}.

Given β_{M_t} and the conjugate prior of Σ_{M_t}, we can sample $\Sigma_{M_t=j}$ from

$$IW\left(R_0 + \sum_{t=1}^{T} \mathbf{I}(M_t = j) \times (Q_t - x_t \times \beta_{M_t})(Q_t - x_t \times \beta_{M_t})',\ v_0 + \sum_{t=1}^{T} \mathbf{I}(M_t = j)\right)$$

for each state $j(= 0$ and $1)$, where $\mathbf{I}(\cdot)$ is an indicator function.

B.1.4 Sampling $(\gamma_{B,M_t}, \lambda_{S,M_t}, \delta_{A,M_t}, \kappa_{B,M_t}, \sigma^2_{B,M_t})$

Next, the parameters in the Bitcoin equation among the measurement equations are sampled in a linear regression framework.

$$B_t|\theta, \mathbf{F}, M_t \sim \mathcal{N}\left(G_t \times \gamma_{B,M_t} + L_t^S \times \lambda_{S,M_t} + L_t^A \times \delta_{A,M_t} + L_t^D \times \kappa_{B,M_t}, \sigma^2_{B,M_t}\right)$$

B.1.5 Sampling P

The transition probability p_{00} is updated by the beta distribution, $Beta(a_0 + n_{00}, b_0 + 1)$, given its conjugate prior and $\mathbf{M}$, where n_{00} is the number of the transitions from state 0 to state 0.

B.1.6 Sampling $\mathbf{M} = \{M_t\}_{t=1}^T$

The final stage is sampling the states. We use the forward and backward recursions of Carter and Kohn (1994). Given the initial condition, $\Pr(M_0 = 0|\theta, \mathbf{F}, \mathbf{Y}_0) = 1$, the forward recursion is the Hamilton filtering to compute the filtered probabilities,

$$\Pr(M_t|\theta, \mathbf{F}, \mathbf{Y}_t) = \frac{f(Y_t|\theta, \mathbf{F}, M_t, \mathbf{Y}_{t-1}) \Pr\left(M_t|\theta, \mathbf{F}, \mathbf{Y}_{t-1}\right)}{f\left(Y_t|\theta, \mathbf{F}, \mathbf{Y}_{t-1}\right)}$$

for all $t = 1, 2, .., T$, where $f(Y_t|\theta, \mathbf{F}, M_t, \mathbf{Y}_{t-1}) = \mathcal{N}(Y_t|H_{M_t} \times F_t,\ \Lambda_{M_t})$ is the conditional likelihood,

$$\Pr\left(M_t = j|\theta, \mathbf{F}, \mathbf{Y}_{t-1}\right) = \sum\nolimits_{i=0}^{1} \Pr(M_t|M_{t-1} = i) \Pr\left(M_{t-1} = i|\theta, \mathbf{F}, \mathbf{Y}_{t-1}\right)$$

is the predictive probability of state, and

$$f\left(Y_t|\theta, \mathbf{F}, \mathbf{Y}_{t-1}\right) = \sum\nolimits_{j=0}^{1} f(Y_t|\theta, \mathbf{F}, M_t = j, \mathbf{Y}_{t-1}) \Pr\left(M_t = j|\theta, \mathbf{F}, \mathbf{Y}_{t-1}\right)$$

is the likelihood density at time t. For the backward recursion, we sample M_T with the probability $\Pr(M_T|\theta, \mathbf{F}, \mathbf{Y}_T)$. Then, using M_{t+1} and the filtered probabilities, M_t for $t = T-1,\ T-2,\ ..,\ 1$ can be sampled from the conditional probability,

$$\Pr(M_t|\theta, \mathbf{F}, \mathbf{Y}_t, M_{t+1}) = \frac{\Pr(M_{t+1}|M_t) \Pr(M_t|\theta, \mathbf{Y}_t)}{\sum_{i=0}^{1} \Pr(M_{t+1}|M_t = i) \Pr(M_t = i|\theta, \mathbf{Y}_t)}$$

in reverse order.

References

Acemoglu, Daren. 2021. "The Bitcoin Fountainhead." *Project Syndicate*.

Adler, Michael, and Bernard Dumas. 1983. "Common Risk Factors in Cryptocurrency." *Journal of Finance*, 38(3): 925–984.

Ahn, Dong-Hyun, Conrad S. Jennifer, and F. Robert Dittmar. 2009. "Basis Assets." *The Review of Financial Studies*, 22(12): 5133–5174.

Auer, Raphael, Marc Farag, Lewrick Ulf, Lovrenc Orazem, and Markus Zoss. 2022. "Banking in the shadow of Bitcoin? The institutional adoption of cryptocurrencies." *BIS working paper*.

Barberis, Nicholas, and Andrei Shleifer. 2003. "Style Investing." *Journal of Financial Economics*, 68: 161–199.

Barberis, Nicholas, Andrei Shleifer, and Jeffrey Wurgler. 2005. "Comovement." *Journal of Financial Economics*, 75(2): 283–317.

Bekaert, G., and Michael S. Urias. 1996. "Diversification, Integration and Emerging Market Closed-End Funds." *Journal of Finance*, 51(3): 835–869.

Bianchi, Daniele, and Mykola Babiak. 2021. "A Factor Moidel for Cryptocurrency Returns." *Available at SSRN 3935934*.

Borri, Nicola. 2019. "Conditional tail-risk in cryptocurrency markets." *Journal of Empirical Finance*, 50: 1–19.

Borri, Nicola, and Paolo Santucci de Magistris. 2021. "Crypto Premium, Higher-Order Moments and Tail Risk." *Available at SSRN 3889169*.

Carter, C., and R. Kohn. 1994. "On Gibbs sampling for state space models." *Biometrika*, 81: 541–53.

Chen, Andrew Y., and Tom Zimmermann. 2022. "Open Source Cross-Sectional Asset Pricing." *Critical Finance Review*, 27(2): 207–264.

Chen, Zhiwu, and Peter J. Knez. 1995. "Measurement of Market Integration and Arbitrage." *Review of Financial Studies*, 8(2): 287–325.

Chib, S. 1995. "Marginal likelihood from the Gibbs output." *Journal of the American Statistical Association*, 90: 1313–1321.

Cho, D. Chinhyung., Chel S. Eun, and Lemma W. Senbet. 1986. "International Asset Pricing Theory: An Empirical Investigation." *Journal of Finance*, 41(2): 313–329.

Cousineau, Denis, and Louis Laurencelle. 2011. "Non-central t distribution and the power of the t test: A rejoinder." *Tutorials in Quantitative Methods for Psychology*, 7(1): 1–4.

Darwin, Charles. 1875. *The Variation of Animals and Plants under Domestication, 2nd Ed.* London: John Maurray.

DeSantis, George. 1995. "Volatility bounds for stochastic discount factors: Tests and implications from international financial markets." *unpublished manuscript.*

Driscoll, Carlos A., David W. Macdonald, and Stephen J. O'Brien. 2009. "From Wild Animals to Domestic Pets, an Evolutionary View of Domestication." *Proceedings of the National Academy of Science*, 106: 9971–9978.

Fama, Eugene F., and Kenneth R. French. 2015. "A Five-Factor Asset Pricing Model." *Journal of Financial Economics*, 116: 1–22.

Fidelity. 2021. "The institutional investor digital assets study."

Gooch, W. Jan. 2010. "Encyclopedic Dictionary of Polymers, 2nd Ed." *Springer*, 1.

Greene, H. William. 1990. *Econometric Analysis.* Macmillan Publishing Company.

Hansen, P. Lars, and Ravi Jagannathan. 1991. "Implications of Security Market Data for Models of Dynamic Economies." *Journal of Political Economy*, 99(2): 225–262.

Hansen, P. Lars, and Ravi Jagannathan. 1997. "Assessing Specification Errors in Stochastic Discount Factor Models." *Journal of Finance*, 52(2): 557–590.

Harrison, Michael J., and David M. Kreps. 1979. "Martingales and Arbitrage in Multiperiod Securities Markets." *Journal of Economic Theory*, 20: 381–408.

Hou, Kewei, Chen Xu, and Lu Zhang. 2015. "Digesting Anomalies: An Investment Approach." *Review of Financial Studies*, 28: 650–705.

Hou, Kewei, Chen Xu, and Lu Zhang. 2020. "Replicating Anomalies." *Review of Financial Studies*, 33: 2019–2133.

Hou, Kewei, Haitao Mo, Chen Xu, and Lu Zhang. 2021. "An Augmented q-factor Model with Expected Growth." *Review of Finance*, 25(1): 1–41.

Huberman, Gur, and Shmuel Kandel. 1987. "Mean-Variance Spanning." *Journal of Finance*, 42: 873–888.

Jagannathan, Ravi, and Zhenyu Wang. 1996. "The Conditional CAPM and the Cross-Section of Expected Returns." *Journal of Finance*, 51(1): 3–53.

Kim, Young Min, and Kyu Ho Kang. 2019. "Likelihood Inference for Dynamic Linear Models with Markov Switching Parameters: On the Efficiency of the Kim Filter." *Econometric Reviews*, 38(10): 1109–1130.

Korajczyk, Robert. 1996. "A Measure of Stock Market Integration for Developed and Emerging Markets." *The World Bank Economic Review*, 10(2): 267–289.

Korajczyk, Robert, and Claude J. Viallet. 1989. "An Empirical Investigationl of International Asset Pricing." *Review of Financial Studies*, 2(4): 353–385.

Krishnan, Marakatha. 1967. "The Moments of a Doubly Noncentral t-Distribution." *Journal of the American Statistical Association*, 32: 278–287.

Lettau, Martin, and Markus Pelger. 2020. "Factors That Fit the Time Series and Cross-Section of Stock Returns." *The Review of Financial Studies*, 33(5): 2274–2325.

Lewellen, Jonathan, Stefan Nagel, and Jay Shanken. 2010. "A skeptical appraisal of asset pricing tests." *Journal of Financial Economics*, 96(2): 175–194.

Liu, Yukun, Aleh Tsyvinski, and Xi Wu. 2022. "Common Risk Factors in Cryptocurrency." *Journal of Finance*, 77(2): 1133–1177.

Liu, Yukun, and Aleh Tsyvinski. 2021. "Risks and Returns of cryptocurrency." *The Review of Financial Studies*, 34: 2689–2727.

OECD. 2022. "Institutionalisation of crypto-assets and DeFi–TradFi interconnectedness."

Omori, Y., S. Chib, N. Shephard, and J. Nakajima. 2007. "Stochastic volatility with leverage: fast and efficient likelihood inference." *Journal of Econometrics*, 140: 425–449.

Park, J.H. 1961. "Moments of the Generalized Rayleigh Distribution." *Quarterly of Applied Mathematics*, 19(1): 45–49.

Price, Edward O. 1984. "Behavioral Aspects of Animal Domestication." *Quarterly Review of Biology*, 59: 1–32.

PwC. 2022. "4th Annual Global Crypto Hedge Fund Report 2022." *PwC Publishing.*

Roubini, Nouriel. 2021. "Bitcoin is not a hedge against tail risk." *Financial Times.*

Street, State. 2021. "Digital assets survey."

Stulz, René M. 1981. "A Model of International Asset Pricing." *Journal of Financial Economics*, 9: 383–406.

Taleb, N. Nassim. 2020. "Bitcoin, Currencies, and Fragility." *Quntitative Finance. Forthcoming.*

Table 1
Variance decomposition of Bitcoin return

	Break Date	G_t	L_t^S	L_t^A	L_t^D	L_t^B	total
Bitcoin	February 21, 2020	(a) Before the break					
		0.010	0.002	0.026	0.007	0.956***	1.000
		(0.016)	(0.003)	(0.037)	(0.012)	(0.052)	
		(b) After the break					
		0.020	0.627***	0.044	0.061	0.249**	1.000
		(0.026)	(0.128)	(0.054)	(0.048)	(0.122)	
MVDA Crypto Market Index	February 21, 2020	(a) Before the break					
		0.004	0.001	0.015	0.006	0.973***	1.000
		(0.008)	(0.002)	(0.018)	(0.010)	(0.027)	
		(b) After the break					
		0.046	0.219**	0.326**	0.116	0.294*	1.000
		(0.055)	(0.096)	(0.186)	(0.080)	(0.192)	

Note: This table presents the posterior means of the variance decomposition before and after the break. The posterior standard errors are in parentheses. The three asterisks, *, ** and ***, correspond to statistical significance with p-value less than 10%, 5%, and 1% respectively.

Table 2
ADL-SV model: Overnight Effect

(a) S&P500

	without FTSE				with FTSE			
	Before the break		After the break		Before the break		After the break	
	mean	S.D.	mean	S.D.	mean	S.D.	mean	S.D.
μ_r	**0.031**	0.007	**0.031**	0.007	**0.032**	0.005	**0.032**	0.005
S_{t-1}^{US}	-0.007	0.011	-0.007	0.011	**-0.061**	0.009	**-0.061**	0.009
R_t^B	-0.001	0.002	**0.020**	0.005	0.000	0.002	**0.011**	0.003
S_t^{UK}					**0.230**	0.011	**0.230**	0.011
μ_h	**-0.070**	0.022	**-0.070**	0.022	**-0.332**	0.097	**-0.332**	0.097
h_{t-1}	**0.970**	0.009	**0.970**	0.009	**0.880**	0.034	**0.880**	0.034
$\log RV_t^B$	0.034	0.022	0.014	0.033	**0.144**	0.045	-0.031	0.051
$\log RV_t^{UK}$					**0.350**	0.105	**0.350**	0.105
σ^2	**0.079**	0.019	**0.079**	0.019	**0.158**	0.045	**0.158**	0.045

(b) Nasdaq

	without FTSE				with FTSE			
	Before the break		After the break		before the break		After the break	
	mean	S.D.	mean	S.D.	mean	S.D.	mean	S.D.
μ_r	**0.070**	0.012	**0.070**	0.012	**0.060**	0.010	**0.060**	0.010
S_{t-1}^{US}	0.030	0.018	0.030	0.018	**-0.067**	0.015	**-0.067**	0.015
R_t^B	-0.001	0.002	**0.045**	0.008	0.000	0.003	**0.032**	0.007
S_t^{UK}					**0.459**	0.017	**0.459**	0.017
μ_h	**-0.076**	0.022	**-0.076**	0.022	**-0.562**	0.116	**-0.562**	0.116
h_{t-1}	**0.696**	0.027	**0.696**	0.027	**0.667**	0.066	**0.667**	0.066
$\log RV_t^B$	0.039	0.028	0.038	0.046	**0.275**	0.066	0.086	0.084
$\log RV_t^{UK}$					**0.859**	0.179	**0.859**	0.179
σ^2	**0.144**	0.034	**0.144**	0.034	**0.320**	0.064	**0.320**	0.064

Note: This table presents the posterior summary of the ADL-SV model on the overnight returns on the U.S. stock indices. *mean* and *S.D.* are the posterior mean and standard deviation, respectively. Bold font indicates that the 95 percent credibility interval does not contain zero.

Table 3
ADL-SV model: Weekend Effect

(a) S&P500

	without FTSE				with FTSE			
	Before the break		After the break		Before the break		After the break	
	mean	S.D.	mean	S.D.	mean	S.D.	mean	S.D.
μ_r	0.018	0.018	0.018	0.018	**0.024**	0.012	**0.024**	0.012
S_{t-1}^{US}	-0.009	0.029	-0.009	0.029	**-0.052**	0.021	**-0.052**	0.021
R_t^B	-0.003	0.004	**0.020**	0.008	0.001	0.003	**0.010**	0.005
S_t^{UK}					**0.282**	0.023	**0.282**	0.023
μ_h	**-0.284**	0.122	**-0.284**	0.122	**-0.772**	0.191	**-0.772**	0.191
h_{t-1}	**0.861**	0.054	**0.861**	0.054	**0.703**	0.070	**0.703**	0.070
$\log RV_t^B$	0.214	0.136	0.061	0.181	**0.520**	0.167	-0.052	0.209
$\log RV_t^{UK}$					**0.780**	0.217	**0.780**	0.217
σ^2	**0.417**	0.172	**0.417**	0.172	**0.473**	0.165	**0.473**	0.165

(b) Nasdaq

	without FTSE				with FTSE			
	Before the break		After the break		before the break		After the break	
	mean	S.D.	mean	S.D.	mean	S.D.	mean	S.D.
μ_r	0.049	0.031	0.049	0.031	0.040	0.023	0.040	0.023
S_{t-1}^{US}	0.062	0.044	0.062	0.044	-0.020	0.033	-0.020	0.033
R_t^B	-0.005	0.006	**0.057**	0.015	0.000	0.005	**0.035**	0.011
S_t^{UK}					**0.531**	0.037	**0.531**	0.037
μ_h	**-0.208**	0.094	**-0.208**	0.094	**-0.651**	0.164	**-0.651**	0.164
h_{t-1}	**0.786**	0.079	**0.786**	0.079	**0.576**	0.097	**0.576**	0.097
$\log RV_t^B$	0.204	0.144	0.119	0.199	**0.428**	0.177	-0.019	0.227
$\log RV_t^{UK}$					**0.937**	0.246	**0.937**	0.246
σ^2	**0.482**	0.206	**0.482**	0.206	**0.484**	0.181	**0.484**	0.181

Note: This table presents the posterior summary of the ADL-SV model on the weekend/holiday returns on the U.S. stock indices. *mean* and *S.D.* are the posterior mean and standard deviation, respectively. Bold font indicates that the 95 percent credibility interval does not contain zero.

Table 4
Summary Statistics of the Data

			# of Assets	μ				σ			
				Mean	Median	Minimum	Maximum	Mean	Median	Min	Max
Pre-Break Period (Jaunary 2017 -Feburary 2020) (T=82)	Basis Assets	Stock Porfolios	25	1.0039	1.0037	0.9998	1.0086	0.0324	0.0330	0.0229	0.0400
		Risk-Free	1	1.0009	1.0009	1.0001	1.0014				
	Individual Assets	Stocks	2,649	1.0040	1.0041	0.9092	1.1126	0.0884	0.0672	0.0147	1.6010
		Cryptoassets	43	1.0796	1.0722	1.0230	1.1889	0.4420	0.3663	0.1814	1.0111
Post-Break Period (March 2020 -March 2022) (T=52)	Basis Assets	Stock Porfolios	25	1.0074	1.0068	1.0044	1.0179	0.0585	0.0589	0.0407	0.0788
		Risk-Free	1	1.0001	1.0000	1.0000	1.0008				
	Individual Assets	Stocks	2,649	1.0107	1.0074	0.9482	1.3390	0.1217	0.0944	0.0328	2.3944
		Cryptoassets	415	1.0664	1.0601	0.9696	1.4418	0.3887	0.3367	0.0290	1.4888

Note: This table summarizes the descriptive statistics of basis assets, which comprises the Fama-French size and book-to-market sorted portfolios and the locally risk-free asset. It also provides the summary statistics of all individual assets available in CRSP and individual cryptoassets with market capitalization greater than 1 million US dollar. μ and σ refer to the mean and standard deviation of biweekly returns on the relevant variable. The returns or payoffs are measured in gross terms. We split the whole time-series into the pre-break period and the post-break period on the basis of the structural break date, February 24, 2020, which is empirically identified in Section 3.

Table 5
Estimation Results of Regressions Between $|\alpha_c^*|$ and Idiosyncratic Risks

(a) Regressand=$|\hat{\alpha}_c^*|$ and Regressor=s_{e_c} $\left(= \widehat{\|e_c\|}\right)$

Before Break	Individual Stocks	2,649	Estimate	0.0057	0.0948	0.4170
			z-value	0.5016	6.0538***	
	Individual Cryptoassets	43	Estimate	-0.0084	0.1791	0.5669
			z-value	-0.3569	2.8867***	
After Break	Individual Stocks	2,649	Estimate	0.0019	0.1096	0.4111
			z-value	2.4240*	11.4644***	
	Individual Cryptoassets	415	Estimate	0.0095	0.1795	0.2969
			z-value	1.1394	5.5080***	

(b) Regressand=$|\hat{\alpha}_c^*|$ and Regressor=s_{ω_c} $\left(= \widehat{\|\omega_c\|}\right)$

		# of Assets		Intercept	Slope	adjusted R^2
Before Break	Individual Stocks	2,649	Estimate	0.0006	0.0945	0.4152
			z-value	0.5190	5.9172***	
	Individual Cryptoassets	43	Estimate	-0.0106	0.1803	0.5747
			z-value	-0.5599	2.9686***	
After Break	Individual Stocks	2,649	Estimate	0.0019	0.1089	0.4108
			z-value	2.4806**	11.4934***	
	Individual Cryptoassets	415	Estimate	0.0096	0.1786	0.2971
			z-value	1.1631	6.6451***	

Note: This table reports the regression results between $|\hat{\alpha}_c^*|$ and $\widehat{\|e_c\|}$ in Panel (a) and those between $|\hat{\alpha}_c^*|$ and $\widehat{\|\omega_c\|}$ in Panel (b). We use the Generalized Method of Moment (GMM) with the Newey-West adjustment for serial correlations with lag 4.

Table 6-A

Estimation Results of Domestication Measures Based on $\|e_c\|$: Before the Break

	Statistics	Mean	Median	S.D.	95% Confidence lower end	95% Confidence upper end	97.5% Confidence lower end	97.5% Confidence upper end	99% Confidence lower end	99% Confidence upper end
Individual Stocks	$s_{e_c} = \widehat{\|e_c\|}$	0.0782	0.0567	0.0763		0.1939		0.2396		0.3244
	$\widehat{\left(\frac{\alpha_c^*}{\|e_c\|}\right)}$	0.0220	0.0259	0.1245	-0.2343	0.2544	-0.2681	0.2888	-0.3262	0.3513
	$\hat{\alpha}_c^*$	0.0007	0.0013	0.0138	-0.0282	0.0245	-0.0376	0.0331	-0.0508	0.0540
	$\widehat{\alpha_c^*\|e_c\|}$	0.0002	0.0000	0.0062	-0.0051	0.0038	-0.0074	0.0069	-0.0110	0.0151

	Statistics	Mean	Median	S.D.	95% Confidence <lower end	95% Confidence Inside	95% Confidence >upper end	97.5% Confidence <lower end	97.5% Confidence Inside	97.5% Confidence >upper end	99% Confidence <lower end	99% Confidence Inside	99% Confidence >upper end
Crypto-Assets	$s_{e_c} = \widehat{\|e_c\|}$	0.4428	0.3739	0.2017		1 (2.33%)	42 (97.67%)		2 (4.65%)	41 (95.35%)		11 (25.58%)	32 (74.42%)
	$\widehat{\left(\frac{\alpha_c^*}{\|e_c\|}\right)}$	0.1559	0.1546	0.0600	0 (0.00%)	41 (95.43%)	2 (4.65%)	0 (0.00%)	43 (100.00%)	0 (0.00%)	0 (0.00%)	43 (100.00%)	0 (0.00%)
	$\hat{\alpha}_c^*$	0.0709	0.0632	0.0476	0 (0.00%)	3 (6.97%)	40 (93.02%)	0 (0.00%)	9 (20.93%)	34 (79.06%)	0 (0.00%)	14 (32.56%)	29 (67.44%)
	$\widehat{\alpha_c^*\|e_c\|}$	0.0385	0.0232	0.0486	0 (0.00%)	1 (2.32%)	42 (97.67%)	0 (0.00%)	4 (9.30%)	39 (90.69%)	0 (0.00%)	11 (25.58%)	32 (74.41%)

Note: This table presents the estimation results of domestication measures associated with $\|e_c\|$ **before** the structural break. The upper part of it reports the cross-sectional distributions of the estimate of the idiosyncratic risk (s_{e_c}) along with the estimates of the candidate domestication measures associated with e_c ($\hat{\alpha}_c^*$, $\widehat{\left(\frac{\alpha_c^*}{\|e_c\|}\right)}$ and $\widehat{\alpha_c^*\|e_c\|}$) of the individual stocks. For each statistic, it report its mean, median and standard deivtation (S.D.) and also the lower and upper-end values of 95%, 97.5% and 99% confidence intervals. Only the upper end value of $s_{e_c} = \widehat{\|e_c\|}$ is reported since it is always positive. These lower and upper end values of each statistic are used as critical values in determining whether a particular cryptoasset is domesticated. The results are presented in the lower part of the table, which reports the number and the percentage of cryptos belonging to each cell (<lower end, inside, >upper end).

Table 6-B

Estimation Results of Domestication Measures Based on $\lVert e_c \rVert$: After the Break

	Statistics	Mean	Median	S.D.	95% Confidence			97.5% Confidence			99% Confidence		
					lower end		upper end	lower end		upper end	lower end		upper end
Individual Stocks	$s_{e_c} = \widehat{\lVert e_c \rVert}$	0.0931	0.0680	0.0909			0.2164			0.2899			0.4328
	$\widehat{\left(\frac{\alpha_c^*}{\lVert e_c \rVert}\right)}$	0.0153	0.0152	0.1662	-0.3035		0.3389	-0.3624		0.3892	-0.4240		0.4390
	$\hat{\alpha}_c^*$	0.0020	0.0010	0.0196	-0.0348		0.0454	-0.0424		0.0600	-0.0553		0.0818
	$\widehat{\alpha_c^* \lVert e_c \rVert}$	0.0006	0.0001	0.0106	-0.0054		0.0087	-0.0072		0.0177	-0.0103		0.0306
	Statistics	Mean	Median	S.D.	95% Confidence			97.5% Confidence			99% Confidence		
					<lower end	Inside	>upper end	<lower end	Inside	>upper end	<lower end	Inside	>upper end
Crypto-Assets	$s_{e_c} = \widehat{\lVert e_c \rVert}$	0.3510	0.2965	0.1885		76 (18.31%)	339 (84.69%)		191 (46.02%)	224 (54.98%)		333 (80.24%)	82 (19.76%)
	$\widehat{\left(\frac{\alpha_c^*}{\lVert e_c \rVert}\right)}$	0.1880	0.1981	0.1530	1 (0.24%)	348 (83.86%)	66 (15.90%)	0 (0.00%)	376 (90.60%)	39 (9.39%)	0 (0.00%)	394 (94.93%)	21 (5.06%)
	$\hat{\alpha}_c^*$	0.0686	0.0572	0.0663	7 (1.69%)	151 (36.39%)	257 (61.93%)	5 (1.20%)	211 (50.84%)	199 (47.95%)	2 (0.48%)	278 (66.99%)	135 (32.53%)
	$\widehat{\alpha_c^* \lVert e_c \rVert}$	0.0306	0.0167	0.0465	14 (3.37%)	118 (28.43%)	283 (48.91%)	11 (2.65%)	203 (48.91%)	201 (48.43%)	9 (2.16%)	285 (68.67%)	121 (29.15%)

Note: This table presents the estimation results of domestication measures associated with $\|e_c\|$ **after** the structural break. The upper part of it reports the cross-sectional distributions of the estimate of the idiosyncratic risk (s_{e_c}) along with the estimates of the candidate domestication measures associated with e_c ($\hat{\alpha}_c^*$, $\widehat{\left(\frac{\alpha_c^*}{\|e_c\|}\right)}$ and $\widehat{\alpha_c^*\|e_c\|}$) of the individual stocks. For each statistic, it reports its mean, median and standard deviation (S.D.) and also the lower and upper-end values of 95%, 97.5% and 99% confidence intervals. Only the upper end value of $s_{e_c} = \widehat{\|e_c\|}$ is reported since it is always positive. These lower and upper end values of each statistic are used as critical values in determining whether a particular cryptoasset is domesticated. The results are presented in the lower part of the table, which reports the number and the percentage of cryptos belonging to each cell (<lower end, inside, >upper end).

Table 7-A

Estimation Results of Domestication Measures Based on $\|\omega_c\|$: Before the Break

Individual Stocks

Statistics	Mean	Median	S.D.	95% Confidence: lower end	95% Confidence: upper end	97.5% Confidence: lower end	97.5% Confidence: upper end	99% Confidence: lower end	99% Confidence: upper end
$s_{w_c} = \widehat{\|\omega_c\|}$	0.0782	0.0567	0.0764		0.1937		0.2396		0.3234
$\widehat{\left(\frac{\alpha_c^*}{\|\omega_c\|}\right)}$	0.0220	0.0258	0.1246	-0.2349	0.2557	-0.2677	0.2886	-0.3281	0.3524
$\hat{\alpha}_c^*$	0.0007	0.0013	0.0138	-0.0282	0.0245	-0.0376	0.0331	-0.0508	0.0540
$\widehat{\alpha_c^*\|\omega_c\|}$	0.0002	0.0000	0.0062	-0.0050	0.0038	-0.0074	0.0069	-0.0110	0.0151

Crypto-Assets

Statistics	Mean	Median	S.D.	95% Confidence: <lower end	95% Confidence: Inside	95% Confidence: >upper end	97.5% Confidence: <lower end	97.5% Confidence: Inside	97.5% Confidence: >upper end	99% Confidence: <lower end	99% Confidence: Inside	99% Confidence: >upper end
$s_{w_c} = \widehat{\|\omega_c\|}$	0.4521	0.3844	0.2017		1 (2.33%)	42 (97.67%)		1 (2.33%)	42 (96.67%)		9 (20.93%)	34 (79.07%)
$\widehat{\left(\frac{\alpha_c^*}{\|\omega_c\|}\right)}$	0.1522	0.1544	0.0584	0 (0.00%)	42 (97.67%)	1 (2.33%)	0 (0.00%)	43 (100.00%)	0 (0.00%)	0 (0.00%)	43 (100.00%)	0 (0.00%)
$\hat{\alpha}_c^*$	0.0759	0.0632	0.0476	0 (0.00%)	3 (6.98%)	40 (93.02%)	0 (0.00%)	9 (20.93%)	34 (79.07%)	0 (0.00%)	14 (32.56%)	29 (67.44%)
$\widehat{\alpha_c^*\|\omega_c\|}$	0.0392	0.0236	0.0491	0 (0.00%)	1 (2.32%)	42 (97.67%)	0 (0.00%)	4 (9.30%)	39 (90.70%)	0 (0.00%)	11 (25.58%)	32 (74.42%)

Note: This table presents the estimation results of domestication measures associated with $\|\omega_c\|$ **before** the structural break. The upper part of it reports the cross-sectional distributions of the estimate of the idiosyncratic risk (s_{ω_c}) along with the estimates of the candidate domestication measures associated with ω_c ($\hat{\alpha}_c^*$, $\widehat{\left(\frac{\alpha_c^*}{\|\omega_c\|}\right)}$ and $\widehat{\alpha_c^*\|\omega_c\|}$) of the individual stocks. For each statistic, it reports its mean, median and standard deviation (S.D.) and also the lower and upper-end values of 95%, 97.5% and 99% confidence intervals. Only the upper end value of $s_{e_c} = \widehat{\|\omega_c\|}$ is reported since it is always positive. These lower and upper end values of each statistic are used as critical values in determining whether a particular cryptoasset is domesticated. The results are presented in the lower part of the table, which reports the number and the percentage of cryptos belonging to each cell (<lower end, inside, >upper end). One thing to notice is that all the results associated with $\hat{\alpha}_c^*$ herein is exactly identical to those in Table 6-A.

Table 7-B

Estimation Results of Domestication Measures Based on $\|\omega_c\|$: After the Break

	Statistics	Mean	Median	S.D.	95% Confidence			97.5% Confidence			99% Confidence		
Individual Stocks					lower end		upper end	lower end		upper end	lower end		upper end
	$s_{\omega_c} = \widehat{\|\omega_c\|}$	0.0934	0.0684	0.0914			0.2192			0.2885			0.4329
	$\widehat{\left(\frac{\alpha_c^*}{\|\omega_c\|}\right)}$	0.0153	0.0149	0.1657	-0.3030		0.3373	-0.3580		0.3897	-0.4263		0.4408
	$\hat{\alpha}_c^*$	0.0020	0.0010	0.0196	-0.0348		0.0454	-0.0424		0.0600	-0.0553		0.0818
	$\widehat{\alpha_c^* \|\omega_c\|}$	0.0006	0.0001	0.0108	-0.0054		0.0089	-0.0072		0.0176	-0.0103		0.0305
	Statistics	Mean	Median	S.D.	95% Confidence			97.5% Confidence			99% Confidence		
Crypto-Assets					<lower end	Inside	>upper end	<lower end	Inside	>upper end	<lower end	Inside	>upper end
	$s_{\omega_c} = \widehat{\|\omega_c\|}$	0.3520	0.2978	0.1895		80 (19.28%)	340 (80.72%)		190 (45.78%)	225 (54.22%)		332 (80.00%)	83 (20.00%)
	$\widehat{\left(\frac{\alpha_c^*}{\|\omega_c\|}\right)}$	0.1873	0.1957	0.1523	1 (0.24%)	352 (84.82%)	62 (14.94%)	0 (0.00%)	377 (90.84%)	38 (9.16%)	0 (0.00%)	395 (95.18%)	20 (4.82%)
	$\hat{\alpha}_c^*$	0.0686	0.0572	0.0663	7 (1.69%)	151 (36.39%)	257 (61.93%)	5 (1.20%)	211 (50.84%)	199 (47.95%)	2 (0.48%)	278 (66.99%)	135 (32.53%)
	$\widehat{\alpha_c^* \|\omega_c\|}$	0.0307	0.0166	0.0466	14 (3.37%)	121 (29.16%)	280 (67.47%)	11 (2.65%)	202 (48.67%)	202 (48.67%)	9 (2.17%)	285 (68.67%)	121 (29.16%)

Note: This table presents the estimation results of domestication measures associated with $\|\omega_c\|$ **after** the structural break. The upper part of it reports the cross-sectional distributions of the estimate of the idiosyncratic risk (s_{ω_c}) along with the estimates of the candidate domestication measures associated with ω_c ($\hat{\alpha}_c^*$, $\widehat{\left(\frac{\alpha_c^*}{\|\omega_c\|}\right)}$ and $\widehat{\alpha_c^* \|\omega_c\|}$) of the individual stocks. For each statistic, it reports its mean, median and standard deviation (S.D.) and also the lower and upper-end values of 95%, 97.5% and 99% confidence intervals. Only the upper end value of $s_{e_c} = \widehat{\|\omega_c\|}$ is reported since it is always positive. These lower and upper end values of each statistic are used as critical values in determining whether a particular cryptoasset is domesticated. The results are presented in the lower part of the table, which reports the number and the percentage of cryptos belonging to each cell (<lower end, inside, >upper end). One thing to notice is that all the results associated with $\hat{\alpha}_c^*$ herein is exactly identical to those in Table 6-B.

Table 8
Correlations Between the GMN-SDF of the Stock Basis Assets and the SDFs of Cryptoassets

# of Clusters	Confidence Level	$\hat{\alpha}_c^*$				$\widehat{\left(\frac{\alpha_c^*}{\|e_c\|}\right)}$		$\widehat{\alpha_c^*\|e_c\|}$		$\widehat{\left(\frac{\alpha_c^*}{\|\omega_c\|}\right)}$		$\widehat{\alpha_c^*\|\omega_c\|}$	
		$m^*_{dc\|\mu^*_m}$	$m^*_{nc\|\mu^*_m}$	m^*_{dc}	m^*_{nc}	$m^*_{dc\|\mu^*_m}$	$m^*_{nc\|\mu^*_m}$	$m^*_{dc\|\mu^*_m}$	$m^*_{nc\|\mu^*_m}$	m^*_{dc}	m^*_{nc}	m^*_{dc}	m^*_{nc}
10	95.0%	0.5433	-0.0101	0.4783	0.0215	0.2018	-0.1517	0.4599	0.0926	0.2094	0.0750	0.3840	0.0420
	97.5%	0.4769	-0.0552	0.3676	-0.0202	0.1605	-0.2599	0.3798	0.1402	0.1124	-0.0065	0.3580	0.1658
	99.0%	0.3735	-0.0595	0.4076	0.1381	0.1136	-0.2770	0.3253	0.0153	0.0654	0.1787	0.3200	-0.1456
11	95.0%	0.5723	-0.0094	0.5083	0.0214	0.2028	-0.2064	0.4835	0.0799	0.2057	-0.0530	0.4460	0.0303
	97.5%	0.4781	-0.0544	0.3678	-0.0346	0.1616	-0.2638	0.3710	0.1474	0.1105	0.0011	0.3192	0.1663
	99.0%	0.3692	-0.0562	0.4002	0.1503	0.1465	-0.2813	0.3033	0.0156	0.0540	0.1777	0.3396	-0.1472
12	95.0%	0.5818	-0.0005	0.5459	0.0254	0.2154	-0.2115	0.4813	0.0780	0.2165	-0.0460	0.4760	0.0307
	97.5%	0.4810	-0.0568	0.4072	-0.0399	0.1984	-0.2615	0.3887	0.1486	0.0993	0.0026	0.3150	0.1667
	99.0%	0.3744	-0.0577	0.4219	0.1441	0.0890	-0.2729	0.3008	0.0133	0.0044	0.0549	0.3462	-0.1527
13	95.0%	0.5882	-0.0468	0.5519	0.0164	0.2531	-0.3384	0.4810	0.0764	0.2420	-0.3218	0.4833	0.0374
	97.5%	0.4749	-0.0721	0.4035	-0.0289	0.1983	-0.2268	0.3890	0.1544	0.0989	0.0353	0.3260	0.1793
	99.0%	0.3702	-0.0540	0.4227	0.1463	0.1184	-0.2724	0.2876	0.0017	0.0217	0.0374	0.3564	0.0016
14	95.0%	0.5666	-0.0461	0.5277	0.0193	0.2545	-0.3821	0.4893	0.0637	0.2425	-0.3512	0.4875	0.0126
	97.5%	0.4550	-0.0508	0.4363	0.0431	0.1885	-0.2318	0.3571	0.1479	0.0777	0.0378	0.3324	0.1576
	99.0%	0.3826	-0.0285	0.4337	0.1788	0.1229	-0.3012	0.2250	-0.0113	0.0220	0.0072	0.2327	0.0311
15	95.0%	0.5632	0.0000	0.5273	0.0270	0.2441	-0.4043	0.4689	0.0487	0.2537	-0.3906	0.4628	-0.0029
	97.5%	0.4921	-0.0522	0.4490	0.0436	0.1829	-0.2159	0.3560	0.1267	0.0629	-0.0878	0.3375	0.1434
	99.0%	0.3844	-0.0004	0.4369	0.1883	0.1250	-0.3032	0.2275	-0.0091	0.0220	0.0004	0.2328	0.0117

Note: This table reports the correlation coefficients between the GMN-SDF of the stock basis assets augmented by the locally risk-free asset (m^*_x) and the SDF of the cryptoassets designated as domesticated assets (and non-domesticated assets) by each domestication measure. $m^*_{dc|\mu^*_m}$ ($m^*_{nc|\mu^*_m}$) is the minimum-norm SDF retrieved from the domesticated cryptos under the restriction that it has the mean same to that of the basis assets, i.e., $E(m^*_{dc|\mu^*_m}) = \mu^*_m$. $m^*_{nc|\mu^*_m}$ is the minimum-norm SDF retrieved from the non-domesticated cryptos under the same restriction. In contrast, m^*_{dc} (m^*_{nc}) is the GMN-SDF retrieved from the domesticated (non-domesticated) cryptos without any restriction on its mean.

Table 9

Bootstrapped Distribution of the Correlations Between the GMN-SDF of Basis Assets and the SDFs of Cryptoassets

Domes. Measure	SDFs of Cryptos		Confidence Level	Domesticated Cryptoassets						Non-domesticated Cryptoassets					
				Mean	Median	S.D.	Min.	Max.	p-val.	Mean	Median	S.D.	Min.	Max.	p-val.
			95.0%	0.440	0.445	0.095	-0.108	0.778	0.000	0.042	0.042	0.108	-0.422	0.483	0.347
$\hat{\alpha_c}^*$	$m^*_{dc\mid\mu^*_m}$	$m^*_{nc\mid\mu^*_m}$	97.5%	0.347	0.352	0.107	-0.167	0.776	0.002	0.037	0.038	0.104	-0.458	0.445	0.359
			99.0%	0.268	0.272	0.111	-0.228	0.651	0.011	-0.014	-0.014	0.102	-0.455	0.443	0.555
			95.0%	0.426	0.431	0.097	-0.110	0.784	0.000	0.070	0.071	0.125	-0.458	0.557	0.289
$\hat{\alpha_c}^*$	m^*_{dc}	m^*_{nc}	97.5%	0.340	0.345	0.110	-0.195	0.749	0.002	0.079	0.080	0.125	-0.467	0.557	0.264
			99.0%	0.271	0.275	0.118	-0.257	0.689	0.015	0.075	0.076	0.127	-0.495	0.611	0.277
			95.0%	0.177	0.179	0.110	-0.319	0.579	0.058	-0.167	-0.167	0.108	-0.601	0.317	0.938
$\widehat{\frac{\alpha_c^*}{\|e_c\|}}$	$m^*_{dc\mid\mu^*_m}$	$m^*_{nc\mid\mu^*_m}$	97.5%	0.126	0.127	0.112	-0.344	0.550	0.132	-0.106	-0.108	0.066	-0.380	0.196	0.943
			99.0%	0.099	0.100	0.113	-0.374	0.538	0.190	-0.160	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*
			95.0%	0.385	0.389	0.102	-0.197	0.739	0.000	0.061	0.061	0.107	-0.382	0.489	0.286
$\widehat{\alpha_c^*\|e_c\|}$	$m^*_{dc\mid\mu^*_m}$	$m^*_{nc\mid\mu^*_m}$	97.5%	0.272	0.275	0.110	-0.203	0.656	0.010	0.064	0.065	0.103	-0.360	0.479	0.267
			99.0%	0.212	0.215	0.113	-0.279	0.606	0.034	0.035	0.036	0.099	-0.449	0.460	0.361
			95.0%	0.175	0.176	0.110	-0.330	0.596	0.060	-0.204	-0.205	0.104	-0.591	0.250	0.973
$\widehat{\frac{\alpha_c^*}{\|\omega_c\|}}$	m^*_{dc}	m^*_{nc}	97.5%	0.124	0.125	0.112	-0.455	0.578	0.135	-0.083	-0.084	0.056	-0.316	0.152	0.928
			99.0%	0.099	0.100	0.113	-0.441	0.557	0.191	-0.098	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*
			95.0%	0.382	0.386	0.101	-0.158	0.723	0.000	0.060	0.061	0.107	-0.420	0.477	0.288
$\widehat{\alpha_c^*\|\omega_c\|}$	m^*_{dc}	m^*_{nc}	97.5%	0.272	0.276	0.110	-0.257	0.661	0.010	0.063	0.064	0.102	-0.387	0.475	0.268
			99.0%	0.212	0.215	0.112	-0.371	0.653	0.034	0.035	0.036	0.099	-0.400	0.401	0.359

Note: This table presents the bootstrapped sampling distribution of correlation between the GMN-SDF retrieved from the stock basis assets augmented by the locally risk-free asset and the relevant SDF retrieved from the cryptos that are designated as domesticated (non-domesticated) assets by each domestication measure. S.D. stands for standard deviation and p-value refers to the p-value of **zero** correlation. With 99% confidence level, $\widehat{\frac{\alpha_c^*}{\|\omega_c\|}}$ ($\widehat{\frac{\alpha_c^*}{\|\omega_c\|}}$) designate only 21 (20) cryptoassets as non-domesticated assets [see Table 6-B (Table 7-B)]. This number is below 25, the number of crypto basis assets pre-specified in bootstrapping. As such, we report the result with when using all of the 21 (22) individual cryptos as crypto basis assets in column, 'Mean,' and write down *n.a.* (not available) in the rest of columns.

Table 10-A

Relation Between the SDF of Domesticated Cryptoassets and the Fama-French 5 Factors/Crypto-Market Factors

Factors	95%			97.5%			99%		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
CMOM	−0.047		0.277	0.779		1.083*	0.783		1.062**
	(0.915)		(0.585)	(0.863)		(0.571)	(0.650)		(0.504)
CSMB	−0.670		−0.641	−0.766**		−0.718*	−1.339***		−1.287**
	(0.450)		(0.538)	(0.340)		(0.426)	(0.411)		(0.600)
CMKT	−0.595		−0.261	−0.589		0.032	−0.927		−0.775
	(0.665)		(0.610)	(0.716)		(0.614)	(0.738)		(0.585)
MRMF		−5.583***	−5.216***		−4.599**	−4.793***		−4.734**	−3.831**
		(1.635)	(1.651)		(2.019)	(1.680)		(2.267)	(1.521)
SMB		−1.356	−1.633		−3.082	−3.619		0.649	0.228
		(2.363)	(2.661)		(2.362)	(2.576)		(3.263)	(3.431)
HML		1.928	2.404		0.460	0.640		−0.073	1.153
		(1.563)	(1.991)		(1.498)	(1.524)		(2.051)	(2.175)
RMW		−2.530	−1.640		−2.843	−2.172		−2.375	−1.044
		(2.611)	(2.913)		(2.305)	(2.332)		(4.033)	(3.766)
CMA		−11.062**	−12.582**		−7.815*	−8.907***		−6.688	−10.352***
		(4.219)	(4.737)		(4.529)	(3.050)		(5.272)	(3.571)
Constant	2.349	19.679***	20.384***	1.593	18.945***	19.506***	2.523*	14.280*	15.930**
	(1.717)	(5.914)	(5.586)	(1.552)	(6.361)	(4.868)	(1.453)	(7.654)	(5.951)
Adjusted R^2	-0.0109	0.1336	0.1073	0.0168	0.1374	0.1603	0.1447	0.0929	0.2224

Note: This table reports the estimation results of the regression of $m^*_{dc|\mu^*_m}$ against the Fama-French five factors coupled with the crypto-market factors proposed by Liu, Tsyvinski and Wu (2022). The standard errors of the estimates are inside the parentheses. In addition, one(*), two(**) and three(***) asterisks denote statistical significance with p-values less than 5%, 2.5% and 1% respectively.

Table 10-B

Relation Between the SDF of Non-Domesticated Cryptoassets and the Fama-French 5 Factors/Crypto-Market Factors

Factors	95%			97.5%			99%		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
CMOM	1.061		1.137	3.098***		3.032***	2.522***		2.534***
	(0.983)		(0.904)	(0.800)		(0.827)	(0.519)		(0.532)
CSMB	−2.628***		−2.527***	−1.736***		−1.748***	−1.827***		−1.852***
	(0.568)		(0.523)	(0.457)		(0.437)	(0.344)		(0.305)
CMKT	−0.726		−0.487	−2.193***		−2.108***	−2.843***		−2.991***
	(0.729)		(0.876)	(0.534)		(0.751)	(0.471)		(0.427)
MRMF		−2.519	−1.741		−0.551	1.520		−1.806	1.493
		(2.602)	(2.529)		(2.829)	(1.751)		(2.119)	(1.181)
SMB		−1.083	−2.512		−4.899	−4.877		−6.584	−6.070**
		(6.941)	(7.258)		(3.663)	(2.940)		(4.006)	(2.520)
HML		1.350	2.643		−1.242	1.552		0.113	3.896***
		(4.765)	(4.417)		(3.034)	(2.903)		(2.951)	(1.439)
RMW		−7.747*	−3.986		−3.308	−3.068		−5.052***	−4.759**
		(4.506)	(3.485)		(2.734)	(3.186)		(1.792)	(2.215)
CMA		6.440	1.621		12.276	4.806		10.141	0.657
		(10.341)	(7.774)		(9.535)	(4.847)		(9.004)	(3.277)
Constant	3.344***	4.576	6.908	1.862**	−1.293	1.907	3.219***	4.189	8.173*
	(0.922)	(12.428)	(11.382)	(0.854)	(8.999)	(4.667)	(0.766)	(8.763)	(4.863)
Adjusted R^2	0.1214	-0.0493	0.0499	0.3899	0.0217	0.3672	0.4904	0.0597	0.4981

Note: This table reports the estimation results of the regression of $m^*_{nc|\mu^*_m}$ against the Fama-French five factors coupled with the crypto-market factors proposed by Liu, Tsyvinski and Wu (2022). The standard errors of the estimates are inside the parentheses. In addition, one(*), two(**) and three(***) asterisks denote statistical significance with p-values less than 5%, 2.5% and 1% respectively.

Table 11-A
Relation Between the SDF of Domesticated Cryptoassets and the q-Five Factors/Crypto-Market Factors

Factors	95%		97.5%		99%	
	(1)	(2)	(3)	(4)	(5)	(6)
CMOM		−0.208		0.486		0.389
		(0.672)		(0.639)		(0.518)
CSMB		−0.486		−0.618*		−1.248**
		(0.407)		(0.357)		(0.505)
CMKT		0.008		0.357		−0.467
		(0.533)		(0.633)		(0.506)
R_MKT	−4.425**	−4.273**	−2.725*	−3.156**	−2.574*	−2.257*
	(1.977)	(2.045)	(1.524)	(1.551)	(1.443)	(1.232)
R_ME	−2.634	−2.905	−2.372	−2.109	−0.006	0.230
	(2.710)	(2.552)	(2.461)	(2.871)	(3.126)	(3.188)
R_IA	−9.062*	−9.538**	−10.285**	−9.234***	−7.068	−8.023**
	(4.561)	(4.196)	(4.465)	(3.254)	(4.438)	(3.660)
R_ROE	0.164	1.641	1.250	1.634	−1.184	1.385
	(4.969)	(5.053)	(3.393)	(3.461)	(4.570)	(3.407)
R_EG	−3.498	−4.665	−0.468	0.789	1.698	−0.641
	(7.094)	(6.673)	(5.056)	(4.944)	(6.396)	(5.506)
Constant	20.485**	21.473**	15.617**	12.855*	10.162	11.694
	(9.672)	(8.107)	(7.276)	(7.320)	(8.808)	(8.804)
Adjusted R^2	0.0526	-0.0015	0.1409	0.1267	0.0437	0.1273

Note: This table reports the estimation results of the regression of $m^*_{dc|\mu^*_m}$ against the q^5 factors of Hou et al. (2021) coupled with the crypto-market factors proposed by Liu, Tsyvinski and Wu (2022). The standard errors of the estimates are inside the parentheses. In addition, one(*), two(**) and three(***) asterisks denote statistical significance with p-values less than 5%, 2.5% and 1% respectively.

Table 11-B

Relation Between the SDF of Non-Domesticated Cryptoassets and the q-Five Factors/Crypto-Market Factors

Factors	95%		97.5%		99%	
	(1)	(2)	(3)	(4)	(5)	(6)
CMOM		-0.162		2.359^{**}		2.025^{***}
		(0.877)		(1.014)		(0.528)
CSMB		-2.387^{***}		-1.771^{***}		-1.778^{***}
		(0.609)		(0.559)		(0.359)
CMKT		-0.069		-1.821^{**}		-2.487^{***}
		(0.874)		(0.733)		(0.602)
R_MKT	1.349	1.765	3.032^{*}	3.499^{*}	1.680	2.814^{**}
	(3.480)	(3.293)	(1.590)	(2.036)	(1.675)	(1.127)
R_ME	-10.144	-10.676^{*}	-9.344^{*}	-7.148	-9.506^{*}	-7.477^{*}
	(7.340)	(6.180)	(5.029)	(4.822)	(4.877)	(3.873)
R_IA	-2.388	-3.811	0.375	-0.850	2.693	-0.262
	(8.704)	(8.427)	(5.192)	(5.186)	(4.804)	(3.449)
R_ROE	-4.138	1.743	2.328	3.727	-0.985	1.441
	(12.001)	(11.575)	(7.455)	(4.953)	(5.888)	(3.065)
R_EG	-3.683	-7.777	-1.876	-4.518	1.931	-3.462
	(16.341)	(15.168)	(10.551)	(9.766)	(8.557)	(5.829)
Constant	19.993	22.426	6.434	7.517	5.130	10.211
	(22.367)	(21.000)	(13.207)	(14.030)	(12.970)	(10.852)
Adjusted R^2	-0.0570	0.0175	0.0019	0.3006	0.0359	0.4170

Note: This table reports the estimation results of the regression of $m^*_{nc|\mu^*_m}$ against the the q^5 factors of Hou et al. (2021) coupled with the crypto-market factors proposed by Liu, Tsyvinski and Wu (2022). The standard errors of the estimates are inside the parentheses. In addition, one(*), two(**) and three(***) asterisks denote statistical significance with p-values less than 5%, 2.5% and 1% respectively.

Table 12-A
Domestication Result Based on RPPCA SDF

Panel (A) : Before Structural Break

					95% Confidence			97.5% Confidence			99% Confidence		
Individual Stocks	Statistics	Mean	Median	S.D.	Lower End		Upper End	Lower End		Upper End	Lower End		Upper End
	$\hat{\alpha}_c$	0.0046	0.0045	0.0123	-0.0181		0.0280	-0.0273		0.0384	-0.0414		0.0577
					95% Confidence			97.5% Confidence			99% Confidence		
Crypto Assets	Statistics	Mean	Median	S.D.	Lower End	Inside	Upper End	Lower End	Inside	Upper End	Lower End	Inside	Upper End
	$\hat{\alpha}_c$	0.0727	0.0648	0.0396	0 (0%)	5 (11.63%)	38 (88.37%)	0 (0.00%)	8 (18.60%)	35 (81.40%)	0 (0.00%)	16 (37.21%)	27 (62.79%)

Panel (B) : After Structural Break

					95% Confidence			97.5% Confidence			99% Confidence		
Individual Stocks	Statistics	Mean	Median	S.D.	Lower End		Upper End	Lower End		Upper End	Lower End		Upper End
	$\hat{\alpha}_c$	0.0087	0.0060	0.0177	-0.0178		0.0503	-0.0251		0.0628	-0.0317		0.0818
					95% Confidence			97.5% Confidence			99% Confidence		
Crypto Assets	Statistics	Mean	Median	S.D.	Lower End	Inside	Upper End	Lower End	Inside	Upper End	Lower End	Inside	Upper End
	$\hat{\alpha}_c$	0.0559	0.0464	0.0565	12 (2.88%)	212 (50.84%)	193 (46.28%)	5 (1.20%)	261 (62.59%)	151 (36.21%)	3 (0.72%)	313 (75.06%)	101 (24.22%)

Note: This table presents the estimation results of domestication measures associated with the implied stochastic discount factor constructed upon the five risk-premium principal components following Lettau and Pelger (2020). The upper (lower) panel reports the result for the before (after) sturcutral break. The cross-sectional distributions of the estimate of $\hat{\alpha}_c^*$ of the individual stocks. It reports its mean, median and standard deviation (S.D.) and also the lower and upper-end values of 95%, 97.5% and 99% confidence intervals.

Table 12-B
Correlations Between the RPPCA SDF and the SDFs of Cryptoassets

# of Clusters	Confidence Level	$\hat{\alpha}_c^*$			
		$m^*_{dc\|\mu^*_m}$	$m^*_{nc\|\mu^*_m}$	m^*_{dc}	m^*_{nc}
10	95.0%	0.2481	0.0285	0.1037	-0.0561
10	97.5%	0.3357	0.0483	0.2268	-0.0948
10	99.0%	0.3285	-0.0758	0.1555	-0.0322
11	95.0%	0.2532	0.0239	0.1313	-0.0697
11	97.5%	0.3326	0.0201	0.2288	-0.1248
11	99.0%	0.3224	-0.0808	0.2764	-0.0312
12	95.0%	0.3621	-0.0059	0.2512	-0.0931
12	97.5%	0.3328	0.0406	0.2535	-0.1186
12	99.0%	0.3165	-0.0675	0.3007	-0.0180
13	95.0%	0.3735	-0.0207	0.3149	-0.0765
13	97.5%	0.2937	0.0454	0.2701	-0.1189
13	99.0%	0.3172	-0.0595	0.2994	-0.1314
14	95.0%	0.2864	-0.0204	0.2446	-0.1342
14	97.5%	0.2739	0.0407	0.2796	-0.1224
14	99.0%	0.3245	-0.0699	0.3001	-0.1414
15	95.0%	0.2955	-0.1014	0.2164	-0.1965
15	97.5%	0.2004	-0.0049	0.1879	-0.1252
15	99.0%	0.3203	-0.0469	0.2864	-0.0214

Note: This table reports the correlation coefficients between implied stochastic discount factor from the five risk-premium principal components based on the 35 anomalys from Chen and Zimmermann (2022) and the SDF of the cryptoassets designated as domesticated assets (and non-domesticated assets) by each domestication measure. $m^*_{dc|\mu^*_m}$ ($m^*_{nc|\mu^*_m}$) is the minimum-norm SDF retrieved from the domesticated cryptos under the restriction that it has the mean same to that of the basis assets, i.e., $E(m^*_{dc|\mu^*_m}) = \mu^*_m$. $m^*_{nc|\mu^*_m}$ is the minimum-norm SDF retrieved from the non-domesticated cryptos under the same restriction. In contrast, m^*_{dc} (m^*_{nc}) is the GMN-SDF retrieved from the domesticated (non-domesticated) cryptos without any restriction on its mean.

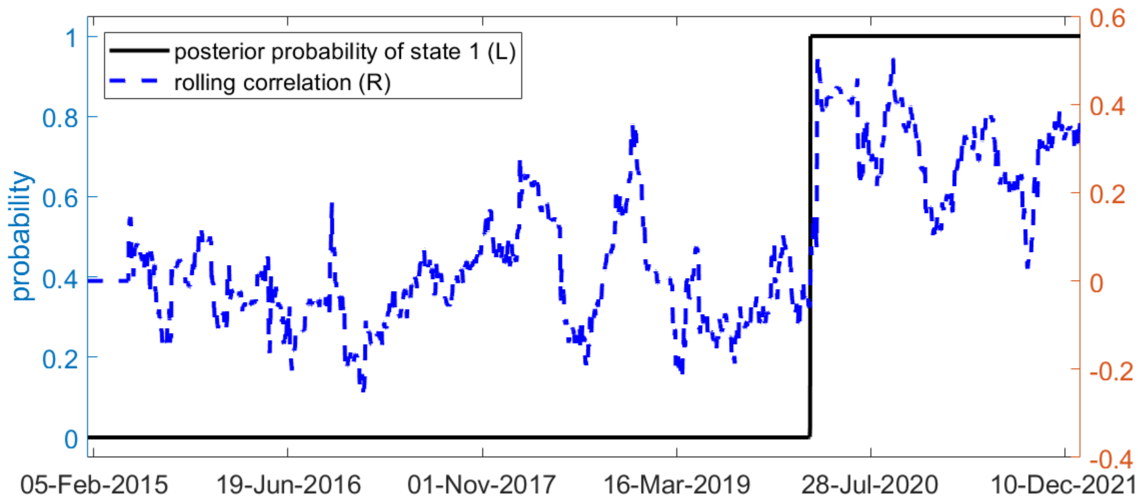
Table 12-C

Bootstrapped Distribution of the Correlations Between the RPPCA SDF and the SDFs of Cryptoassets

Domes. Measure	SDFs of Cryptos	Confidence Level	Domesticated Cryptoassets						Non-Domesticated Cryptoassets					
			Mean	Median	S.D.	Min	Max	p-value	Mean	Median	S.D.	Min	Max	p-value
		95.0%	0.2741	0.2756	0.1130	-0.2569	0.6450	0.0092	-0.0299	-0.0305	0.1022	-0.4125	0.3822	0.6199
α_c^*	$m^*_{c\|\mu^*_m}$	97.5%	0.2647	0.2685	0.1125	-0.2006	0.6374	0.0116	-0.0598	-0.0593	0.0922	-0.4196	0.4105	0.7414
		99.0%	0.2129	0.2159	0.1138	-0.2473	0.5644	0.0358	-0.1100	-0.1109	0.0837	-0.4060	0.2021	0.9044
		95.0%	0.2576	0.2616	0.1183	-0.2299	0.6377	0.0168	-0.0222	-0.0195	0.1282	-0.5288	0.4828	0.5666
α_c^*	m_c^*	97.5%	0.2482	0.2527	0.1183	-0.2619	0.6449	0.0229	-0.0089	-0.0093	0.1224	-0.4850	0.4461	0.5314
		99.0%	0.1899	0.1925	0.1228	-0.3334	0.6197	0.0637	0.0018	0.0032	0.1225	-0.4087	0.4186	0.4913

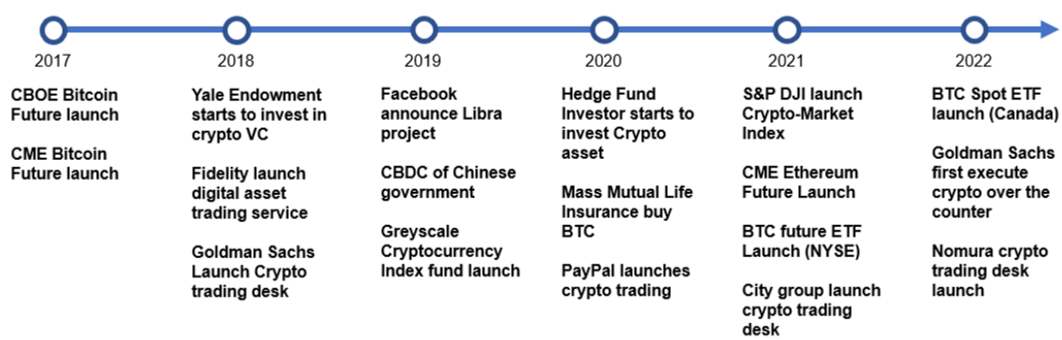
Note: This table presents the bootstrapped sampling distribution of correlation between the implied stochastic discount factor from the five risk-premium principal components based on the 35 anomalys from Chen and Zimmermann (2022) and the relevant SDF retrieved from the cryptos that are designated as domesticated (non-domesticated) assets by domestication measure $\hat{\alpha}_c{}^*$. S.D. stands for standard deviation and p-value refers to the p-value of **zero** correlation.

Figure 1
Posterior probability of state 1



Note: This figure plots the posterior probability of state 1 along with the rolling correlations between the Bitcoin and S&P500 returns, where the window size is 100 business days.

Figure 2
Time line of major events in the crypto market



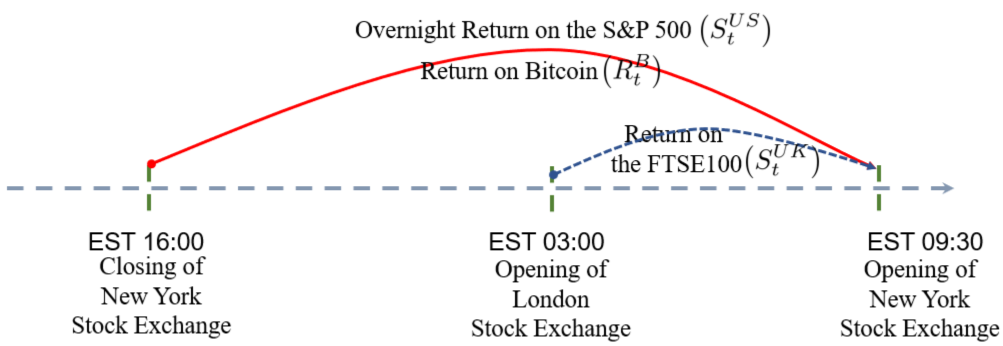
Note: This figure summarizes the major events that have taken place in the crypto market. The institutionalization begins with the inception of futures on Bitcoin in major derivative exchanges such as CBOE and CME.

Table A.1
Simulation Results on Testing the Unbiasedness of Statistics

β_{c0}	σ_{e_c}	Parameter	α_c^*	$\|e\|$	$\frac{\alpha_c^*}{\|e\|}$	$\alpha_c^*\|e\|$	$\|\omega_c\|$	$\frac{\alpha_c^*}{\|\omega_c\|}$	$\alpha_c^*\|\omega_c\|$
0.01	0.01	True	0.0628	0.0100	6.2765	0.0006	0.0100	6.2748	0.0006
		Simulation	0.0628	0.0100	6.2776	0.0006	0.0100	6.2769	0.0006
0.01	0.10	True	0.0628	0.1000	0.6276	0.0006	0.0100	0.6276	0.0063
		Simulation	0.0627	0.1000	0.6274	0.0006	0.0100	0.6275	0.0063
0.01	0.50	True	0.0628	0.5000	0.1255	0.0314	0.5000	0.1255	0.0314
		Simulation	0.0626	0.5000	0.1253	0.0313	0.5000	0.1253	0.0313
0.01	1.00	True	0.0628	1.0000	0.0628	0.0628	1.0000	0.0628	0.0628
		Simulation	0.0624	1.0001	0.0625	0.0624	1.0000	0.0625	0.0624
0.10	0.01	True	0.1521	0.0100	15.2090	0.0015	0.0103	14.8085	0.0016
		Simulation	0.1521	0.0100	15.2126	0.0015	0.0103	14.7970	0.0016
0.10	0.10	True	0.1521	0.1000	1.5209	0.0152	0.1000	1.5205	0.0152
		Simulation	0.1521	0.1000	1.5211	0.0152	0.1000	1.5209	0.0152
0.10	0.50	True	0.1521	0.5000	0.3042	0.0760	0.5000	0.3043	0.0760
		Simulation	0.1523	0.5000	0.3046	0.0761	0.4998	0.3046	0.0761
0.10	1.00	True	0.1521	1.0000	0.1521	0.1521	1.0000	0.1521	0.1521
		Simulation	0.1520	1.0000	0.1520	0.1520	0.9999	0.1520	0.1519
0.50	0.01	True	0.5491	0.0100	54.9091	0.0055	0.0154	35.6625	0.0085
		Simulation	0.5491	0.0100	54.9234	0.0055	0.0155	36.1650	0.0085
0.50	0.10	True	0.5491	0.1000	5.4909	0.0549	0.1007	5.4537	0.0553
		Simulation	0.5491	0.1000	5.4909	0.0549	0.1007	5.4528	0.0053
0.50	0.50	True	0.5491	0.5000	1.0982	0.2745	0.5001	1.0979	0.2746
		Simulation	0.5494	0.5001	1.0986	0.2747	0.5002	1.0984	0.2748
0.50	1.00	True	0.5491	1.0000	0.5491	0.5491	1.0001	0.5491	0.5491
		Simulation	0.5494	1.0001	0.5493	0.5495	1.0001	0.5494	0.5495

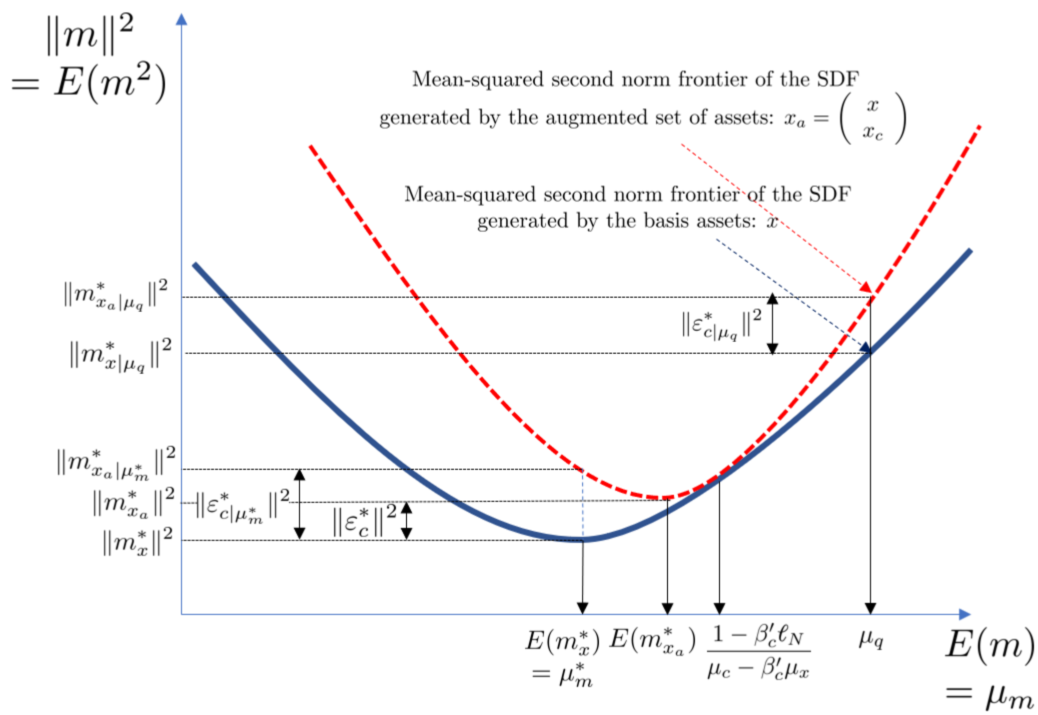
Note: This table summarizes the simulation results on testing the unbiasedness of relevant statistics. We report the mean value of the statistics for each combination of β_{c0} and σ_{e_c} based on 500,000 simulations. In each round of simulation, we compute $\hat{\alpha}_c^*$, which is the unbiased estimator proposed in Proposition 6. Similarly we compute s_{e_c}, $\widehat{\frac{\alpha_c^*}{\|e_c\|}}$ and $\widehat{\alpha_c^*\|e_c\|}$ in Proposition 7-A, which are the unbiased estimators of $\|e_c\|$, $\frac{\alpha_c^*}{\|e_c\|}$ and $\alpha_c^*\|e_c\|$ respectively. In contrast, we compute s_{ω_c}, $\widehat{\frac{\alpha_c^*}{\|\omega_c\|}}$ and $\widehat{\alpha_c^*\|\omega_c\|}$, the pseudo-unbiased estimators of $\|\omega_c\|$, $\frac{\alpha_c^*}{\|\omega_c\|}$ and $\alpha_c^*\|\omega_c\|$ in Proposition 7-B. Here The basis assets employed in simulations are the post-break returns on the Fama-French twenty-five portfolios and the locally risk-free asset. In each round of simulation, we set the estimated post-break betas ($\hat{\beta}_c$) of Bitcoin as the hypothetical true $\hat{\beta}_c$ and simulate x_c given β_{c0}, β_c and σ_{e_c} using the random normal generator.

Figure 3
Time line of the overnight effect



Note: This figure displays the time line used in the analysis of the overnight effect. New York Stock Exchange (NYSE) closes at EST (Eastern Standard Time) 16:00 and opens at EST 09:30 on the following business day. S_t^{US} is the overnight (after-hour) returns on the S&P500 and the Nasdaq. S_t^{UK} is the return on the FTSE100 index from EST 03:00 (when the London Stock Exchange opens) to EST 09:00, 30 mininutes before the opening of the NYSE. R_t^B stands for the return on Bitcoin from EST 16:00 to EST 09:00.

Figure 4
The Mean-Squared Second Norm Bounds for the SDFs



Note: This figure compares the mean-squared second norm bounds for the SDF generated by x (the set of stock basis assets) and x_a (the augmented set of assets). Each of the two bounds is parabolic and they are tangent at $E(m) = \frac{1-\beta_c'\ell_N}{\mu_c-\beta_c'\mu_x}$. The mean and squared second norm of the global minimum-norm SDF generated by the basis assets are μ^*_m and $\|m^*_x\|$ respectively. Similarly, the corresponding moments of the global minimum-norm SDF generated by the augmented set of assets are $E(m^*_{x_a})$ and $\|m^*_{x_a}\|$ respectively. ε^*_c is an orthogonal expansion of m^*_x which leads to $m^*_{x_a}$ whereas $\varepsilon^*_{c|\mu^*_m}$ is an orthogonal extensions of m^*_x which yields $m^*_{x_a|\mu^*_m}$.