UNIVERSITY OF CALIFORNIA

Los Angeles

Essays in Econometrics and Empirical Asset Pricing

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Economics

by

Adam Baybutt

2024

ABSTRACT OF THE DISSERTATION

Essays in Econometrics and Empirical Asset Pricing

by

Adam Baybutt
Doctor of Philosophy in Economics
University of California, Los Angeles, 2024
Professor Denis Nikolaye Chetverikov, Chair

The first and last chapter of this dissertation are devoted to the econometric theory of two unrelated topics. The second chapter covers an empirical study of the novel model in the first chapter.

The first chapter studies novel estimation procedures with supporting econometric theory for a dynamic latent-factor model with high-dimensional asset characteristics, that is, the number of characteristics is on the order of the sample size. Utilizing the Double Selection Lasso estimator, our procedure employs regularization to eliminate characteristics with low signal-to-noise ratios yet maintains asymptotically valid inference for asset pricing tests.

The second chapter studies the dynamics of crypto asset returns through the lens of factor models, and in particular compare the out of sample pricing ability of our novel factor model against relevant benchmarks. We were motivated to develop our new method given, in the setting of crypto asset returns, there are a limited number of tradable assets and years of data as well as a rich set of available asset characteristics. In an additionally novel empirical panel, we find the new estimator obtains comparable out-of-sample pricing ability and risk-adjusted returns to benchmark methods. We provide an inference procedure for

measuring the risk premium of an observable nontradable factor, and employ this to find that the inflation-mimicking portfolio in the crypto asset class has positive risk compensation. Finally, specifying a factor model with nonparametric loadings and factors, we utilize recent methods in deep learning to maximize out-of-sample risk-adjusted returns in an hourly panel, which yields economically significant alphas even after a detailed accounting of transaction costs.

The third chapter (coauthored with Manu Navjeevan) studies a novel estimator for the conditional average treatment effect (CATE) with a doubly-robust inference procedure. Plausible identification of CATEs can rely on controlling for a large number of variables to account for confounding factors. In these high-dimensional settings, estimation of the CATE requires estimating first-stage models whose consistency relies on correctly specifying their parametric forms. While doubly-robust estimators of the CATE exist, inference procedures based on the second-stage CATE estimator are not doubly-robust. Using the popular augmented inverse propensity weighting signal, we propose an estimator for the CATE whose resulting Wald-type confidence intervals are doubly-robust. We assume a logistic model for the propensity score and a linear model for the outcome regression, and estimate the parameters of these models using an ℓ_1 (Lasso) penalty to address the high-dimensional covariates. Inference based on this estimator remains valid even if one of the logistic propensity score or linear outcome regression models are misspecified. To our knowledge, we are the first paper to develop doubly-robust pointwise and uniform inference on an infinite dimensional target parameter after high dimensional nuisance model estimation.

The dissertation of Adam Baybutt is approved.

Valentine P. Haddad

Andres Santos

Pierre-Olivier Weill

Denis Nikolaye Chetverikov, Committee Chair

University of California, Los Angeles

2024

To my wife Rachel,

for her love, strength, and wisdom;

and, to my mother Tamara and father Paul,

for their love, care, and guidance.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am grateful to Denis Chetverikov for guidance, support, and uniquely valuable mathematical insights.

I thank the following people for valuable comments and suggestions: Jesper Böjeryd, Mikhail Chernov, Valentin Haddad, Basil Halperin, Jinyong Hahn, Niklas Henke, Zhipeng Liao, Manu Navjeevan, Matthew Reid, Andres Santos, Pierre-Olivier Weill, and participants in UCLA's Econometrics Proseminar.

I thank the following individuals for their research assistance: Jacob Brophy, Chuhan Guo, and Tiffany Zho.

Thanks to Pascal Michaillat for formatting. Thank you to Coin Metrics, CoinMarketcap, and Glassnode for providing academic research discounts on data purchases.

VITA

2008–2012	B.S. and M.S. in Biomedical Engineering, University of Southern California, Los Angeles, California.
2014–2016	Program Manager and Biostatistician, Cepheid, Sunnyvale, California.
2016–2018	Research Associate, Massachusetts Institute of Technology, Cambridge, Massachusetts.
2019	Research Fellow, Y Combinator Research, Oakland, California.
2020	M.A. in Economics, Department of Economics, UCLA.
2021	Investment Strategy PhD Intern, Dimensional Fund Advisors, Santa Monica, California.
2022	Quantitative Researcher, Krypton Labs, Los Angeles, California.
2019–2024	Teaching Assistant, Department of Economics, UCLA.

CHAPTER 1

Econometric Theory for a Dynamic Latent-Factor Model with High-Dimensional Asset Characteristics

1.1 Introduction

Factor models are the cornerstone cross-sectional asset pricing model. We develop novel estimation procedures with supporting econometric theory for a dynamic latent-factor model with high-dimensional asset characteristics.

We assume a statistical model where asset excess returns $r_{i,t+1} \in \mathbb{R}$ are a function of common time-varying latent (unobserved) factors, $f_{t+1} \in \mathbb{R}^k$, as dictated by time-varying asset-specific factor loadings $\beta_{i,t} \in \mathbb{R}^k$, that is, for all assets $i \in \{1, \dots, N\}$ and time $t \in \{1, \dots, T\}$

$$\begin{aligned} r_{i,t+1} &= \alpha_{i,t} + \beta_{i,t}^\top f_{t+1} + \epsilon_{i,t+1}^r \\ \beta_{i,t} &= \Gamma_\beta^\top z_{i,t} + \epsilon_{i,t}^\beta \end{aligned} \tag{1.1}$$

where $\alpha_{i,t} \in \mathbb{R}$ are average pricing errors of the factor model; $\epsilon_{i,t+1}^r \in \mathbb{R}$ are uncorrelated idiosyncratic errors, i.e., $\mathbb{E}_t[\epsilon_{i,t+1}^r f_{t+1}] = 0$; $\Gamma_\beta \in \mathbb{R}^{p \times k}$ is a static latent loading parameter; and, $z_{i,t} \in \mathbb{R}^p$ are time-varying asset-specific characteristics where p is high-dimensional, i.e., on the order of N and T. Crucially, we follow the established practice in the literature of assuming the number of factors k is low-dimensional, i.e., $N, T, p >> k \in \{1, 2, 3, \dots\}$. This model has theoretical underpinnings motivated by a structural model for asset excess returns or by the assumption of no arbitrage, as discussed in our literature review.

The main contribution of this research is to develop new estimation and inference pro-

cedures, termed the the Double Selection Lasso Factor Model (DSLFM), to fit the latent factors f_{t+1} and loadings Γ_β in (1.1) and to conduct standard asset pricing tests under the novel setting of high-dimensional asset characteristics.

The DSLFM remains consistent with the equilibrium asset pricing principle that risk premia are solely determined by risk exposures and specifies a linear loading mapping Γ_β between characteristics and dynamic factor loadings $\beta_{i,t}$. We have two novel assumptions for Γ_β. First, we develop estimation procedures and large-sample theory that allows $p, T, N \to \infty$. Given our focus is studying the cross-section of crypto assets, this assumption is particularly relevant given the numerous asset characteristics available, as previously discussed, as well as the existence of only a small number of tradable assets and years of relevant data such that p, T, N are of similar order. Second, we assume exact row sparsity in Γ_β; that is, only a small number of the p asset characteristics determine the content of the factor loading, which matches empirical findings in cross-sectional asset pricing (Babiak and Bianchi (2021) and Bianchi, Guidolin, and Pedio (2022)). These novel assumptions within a dynamic latent-factor model require novel estimation procedures and supporting asymptotic theory.

The DSLFM aims to jointly and consistently estimate the loading matrix Γ_β and latent factors f_{t+1}. If we were to utilize the mean squared error (MSE) objective function to minimize over the $p-$dimensional choice vector $\Gamma_\beta f_{t+1}$, for each t, the mean-squared error of $\sum_i (r_{i,t+1} - z_{i,t}^\top \Gamma_\beta f_{t+1})^2$, we will have not only a noisily estimated design matrix when $p \sim N$, (or, at worst, a nonsingular design matrix when $p > N$) but also a non-convex objective function given the interaction between minimization arguments f_{t+1} and Γ_β. The next logical step would be to introduce sparsity in Γ_β, which would amount to adding a regularization parameter to the aforementioned objective function to combat the curse of dimensionality from $z_{i,t}$. However, although potentially helpful for minimizing MSE by decreasing the variance of the estimator, this regularization introduces a bias in estimation, which would lead to invalid asymptotic inference for asset pricing tests, defeating a goal of this research and, more broadly, the purpose of factor models in this field.

We therefore adapt for our purpose the Double Selection Lasso (DSL) estimator developed by Belloni, Chernozhukov, and Hansen (2014). The key insight from their work was to introduce an orthogonality wherein, assuming Γ_β is row sparse, the regularization bias from the LASSO first-stage estimation does not pass through to the target parameter of interest when conducting inference. We first estimate for each time period t and each characteristic j the scalar $\Gamma_{\beta,j}^\top f_{t+1}$ using DSL; then, stacking these estimates into a $T \times p$ matrix, we use PCA to obtain separate estimates for latent loadings $\widehat{\Gamma}_\beta$ and factors $\{\widehat{f}_{t+1}\}_{t=1}^T$; and, finally, we soft-threshold $\widehat{\Gamma}_\beta$ to set numerous rows to zero given the assumption of sparsity for Γ_β.

This procedure has several additional benefits. Given the period-by-period cross-sectional regression—mirroring Fama-MacBeth regressions–our estimation procedure accommodates unbalanced panels. This first DSL step does require running $T \times p$ cross-sectional regressions; however, this can be done in parallel and is on the order of minutes in practice as each regression is computationally fast. In the second step, the high dimensionality of the PCA procedure, given we have a $p \times T$ matrix, is adapted from the established theory for $N \times T$ excess return matrices (Bai (2003)). The final soft-thresholding step exploits the sparsity in Γ_β to remove noise from characteristics with low signal-to-noise ratios. [1]

Under standard DSL assumptions adapted to our setting (Belloni, Chernozhukov, and Hansen 2014), high-dimensional PCA assumptions (Bai 2003), and assuming we observe the true number of latent factors (Bai and Ng 2002), we develop the asymptotic consistency of the latent factors $\widehat{f}_{t+1}$ and loading matrix $\check{\Gamma}_\beta$ for the latent factors f_{t+1} and loadings Γ_β, respectively. Monte Carlo simulations corroborate with finite-sample evidence that the performance of the DSLFM is comparable to or surpasses relevant benchmarks. As is standard in this setting, without further restrictions outlined in Bai and Ng (2013), F^0 and Γ_β^0 are not separately identifiable; hence, the $k \times k$ invertible matrix transformation H appears in

[1] Finally, just as DSL laid the groundwork for the more general Debiased Machine Learning (DML) theory, this work sets up future research to extend the framework with a semi-parametric specification to utilize the rich set of available machine learning estimators that have been shown to handle well the nonlinearities in cross-sectional asset returns (Gu, Kelly, and Xiu (2020)).

each asymptotic result. However, in many cases, knowing F^0H is equivalent to knowing F^0; for example, using the regressor F^0 will give the same predictions as using the regressor F^0H given they have the same column space. Similarly, the target parameter in the coming inference result is rotation invariant to H.

To show the generality of these estimation procedures, we enrich our model—with one of several possible extensions—to address a common question in asset pricing research. We ask whether an observable, nontradable factor $g_{t+1} \in \mathbb{R}$ carries a risk premium: compensation for exposure to the risk factor holding constant exposure to all other sources of risk, i.e., variation with other factors. In the subsequent empirical applications, we investigate a common hypothesis for the crypto asset class: exposure to inflation offers crypto investors a positive risk premium.

Following a recent approach in the literature (Giglio and Xiu (2021) and Giglio, Xiu, and Zhang (2021)), we assume the "true" latent factors f_{t+1} can be decomposed into the latent-factor risk premia $\gamma \in \mathbb{R}^k$ and latent-factor innovations $v_{t+1} \in \mathbb{R}^k$, that is, $f_{t+1} \coloneqq \gamma + v_{t+1}$. Then, we specify the observable factor g_{t+1} as potentially linearly correlated with the latent factors through

$$g_{t+1} = \eta v_{t+1} + \epsilon^g_{t+1},$$

where $\eta \in \mathbb{R}^k$ is an unknown parameter mapping the relation between the latent-factor innovations and the observable factors, and $\epsilon^g_{t+1} \in \mathbb{R}$ is measurement error in g_{t+1}.

The risk premium of an observable factor—our target parameter in this extension—is defined to be the expected excess return of a portfolio with loading (i.e., beta) of 1 with respect to this factor in g_{t+1} and zero loadings on all other factors; in this model, that parameter is $\gamma_g \coloneqq \eta^\top \gamma = \eta^{0\top} H H^{-1} \gamma^0 = \eta^{0\top} \gamma^0$, which utilizes the rotation invariant result of Giglio and Xiu (2021).

We thus extend with our estimation procedure for γ_g in a dynamic latent factor model with high-dimensional characteristics the estimation procedure of Giglio and Xiu (2021),

which is for a static latent factor model. We additionally develop our estimator's large-sample distribution and variance to conduct asset pricing tests on the sign of the observable factor risk premium.

1.2 Literature Review

This paper builds on the literature of the econometric theory of high-dimensional factor models. Giglio, Kelly, and Xiu (2022) provide an excellent review of recent machine-learning based factor model applications and relevant econometric theory, including the common asymptotic frameworks of fixed N and $T \to \infty$, fixed T and $N \to \infty$, and $T, N \to \infty$. In short, this paper extends this last asymptotic framework to be high-dimensional on the new dimension of the number of asset characteristics, i.e., $p, T, N \to \infty$.

Starting from either a structural model for asset excess returns in the style of the Capital Asset Pricing Model (Sharpe 1964), or the assumption of no arbitrage, as in Arbitrage Pricing Theory (Ross 1976), a stochastic discount factor $m_{t+1} \in \mathbb{R}$ exists and an Euler equation, termed the Law of One Price, holds for asset excess returns $r_{i,t+1} \in \mathbb{R}$ for assets $i \in \{1, 2, \ldots, N\}$ in time periods $t \in \{1, 2, \ldots, T\}$

$$\mathbb{E}_t[m_{t+1} r_{i,t+1}] = 0,$$

which by the definition of the variance and covariance operators,

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

As discussed in Section 1.1 we directly assume the statistical model

$$r_{i,t+1} = \alpha_{i,t} + \beta_{i,t}^\top f_{t+1} + \epsilon_{i,t+1}^r.$$

To map this model to its theoretical underpinnings in the Law of One Price, one can assume for all i and t: mean zero unobserved idiosyncratic errors $\mathbb{E}_t[\epsilon_{i,t+1}^r] = 0$, uncorrelated errors

$\mathbb{E}_t[\epsilon^r_{i,t+1} f_{t+1}] = 0$, the price of risk associated with the factors to be defined as $\lambda_t \coloneqq \mathbb{E}_t[f_{t+1}]$, and zero average pricing errors $\alpha_{i,t} = 0$ (Cochrane 2009). The factor model posits that asset excess returns $r_{i,t+1}$ signify compensation for asset-specific, time-varying exposure $\beta_{i,t} \in \mathbb{R}^k$ to systematic risk factors $f_{t+1} \in \mathbb{R}^k$.

The classic factor model is a static loading observable factor model—in the style of Fama and French—where pricing errors and static factor loadings β_i in (1.1) using exogenously-defined factors are estimated via the "Fama-MacBeth" two-step procedure (Fama and MacBeth 1973), which has rich supporting inferential theory (Shanken 1992). This procedure relies on ex ante declaration of observable factors f_{t+1} formed as a convex combination (i.e., a portfolio) of sorted asset returns r_{t+1} based on asset characteristics $z_t \in \mathbb{R}^p$. This is likely an incomplete model of the relationship between z_t and r_{t+1} and prone to overfit.

Recognizing the proliferation of factors, a "Factor Zoo" (Cochrane 2011), explaining the cross section of expected returns, Feng, Giglio, and Xiu (2020) propose the use of Double Selection Lasso (Belloni, Chernozhukov, and Hansen 2014) combined with two-pass Fama-MacBeth regressions to evaluate the contribution of a new factor, g_{t+1}, explaining asset expected returns above and beyond an existing high-dimensional set of factors. However, as the recent empirical literature has shown (e.g., Kelly, Pruitt, and Su (2019); Chen, Pelger, and Zhu (2020)), allowing the data to construct the relevant latent factors offers superior explanatory and predictive power as compared to using a set of selected observable factors from the literature.

Factor models with latent factors have been a focal point since the development of APT (Ross 1976) and early empirical efforts (Chamberlain 1983). PCA is the common estimation framework. Bai (2003) develops the core theory for PCA estimation and inference under joint $N, T \to \infty$ high-dimensional asymptotics, with Bai and Ng (2002) introducing a novel Information Criterion penalization to ascertain the true number of latent factors. Assumptions for the joint identification of factor loadings and factors are summarized in Bai and Ng (2013). This manuscript leverages these contributions yet employs instead a dynamic

latent-factor model with $p, N, T \to \infty$ high-dimensional asymptotics. In the asset pricing context, the static model can still perform well for describing portfolios over time given the dynamics are captured by the dynamic factors; however, this has not performed as well in describing individual asset returns, as noted by Ang, Liu, and Schwarz (2009). Although this model allows the data to statistically inform the factor structure, it fails to incorporate rich asset characteristic data and it assumes a static factor loading that maps systematic risks to excess returns.

A recent and significant methodological advance, instrumented PCA (IPCA) from Kelly, Pruitt, and Su (2019), utilizes asset characteristics in a linear model of dynamic factor loadings to parameterize the more general semi-parametric method studied in Connor and Linton (2007). That is,

$$\beta_{i,t} = \Gamma_\beta^\top z_{i,t} + \epsilon_{i,t}^\beta$$

where $\Gamma_\beta \in \mathbb{R}^{p \times k}$ is a loading matrix mapping asset characteristics to the factor loading. IPCA has several benefits, including compressing the $N \times T$ factor loading matrix β to a lower dimensional $p \times k$ loading matrix Γ_β, as well as specifying a time-varying relationship $\beta_{i,t}$ between characteristics and returns, which as stated previously appears to be the empirical reality in crypto cross-sectional asset pricing. In a separate theory paper, the authors develop the asymptotic distributional theory—following $N, T \to \infty$ asymptotics from Bai (2003)—for the factor realizations and loadings under quite general identifying restrictions on loadings and factors (Kelly, Pruitt, and Su 2020).

The IPCA procedure benefits from the following: the efficiency gains from using asset characteristics for estimating the latent factors and their loadings; accommodation of unbalanced panels; maintaining an expected return factor model structure to ascertain the economic relationships among factors and assets via the observable characteristics; and, a parametric model with inference procedures for asset pricing tests. The IPCA estimation procedure, however, is not possible under high-dimensional asset characteristics (i.e., $p > T, N$), or, if regularization is used, produces biased inference for asset pricing tests.

Giglio and Xiu (2021) develop a three-step procedure combining estimation of latent-factor model via PCA with standard two-pass regressions to recover the risk premium of an observable nontradable factor $g_{t+1}\mathbb{R}$, which are potentially correlated with the latent factors:

$$g_{t+1} = \eta^\top v_{t+1} + \epsilon^g_{t+1}$$

where v_{t+1} are (mean-zero) latent-factor innovations (i.e., $f_{t+1} = \gamma + v_{t+1}$); $\eta \in \mathbb{R}^k$ is a linear mapping of the true latent factors to the observed factors; and, ϵ^g_{t+1} is measurement error. This allows the observable factors to either be some component of the latent factors or just correlated with v_{t+1} and therefore still carry a risk premium. The target parameter is the risk premium associated with the observable factors $\gamma_g \coloneqq \eta\gamma$.

To demonstrate the generality of the DSLFM, we extend our estimation procedure by adding the procedure of Giglio and Xiu (2021) to address this classic asset pricing test of the recovering the risk premium of an observable factor. The DSLFM theory extends Giglio and Xiu (2021) by incorporating not only dynamic factor loadings, but also high-dimensional asset characteristics.

More recent literature has incorporated a wide array of machine learning-based estimation approaches within the factor model structure. Gu, Kelly, and Xiu (2020) study a set of machine learning estimation procedures for measuring the equity risk premium to find that deep learning methods outperform out-of-sample. Gu, Kelly, and Xiu (2021) develop a factor model in the structure of IPCA, but allows for non-linear mappings to the factor loadings and the factors through two feed-forward neural networks. Although, in practice they only use linear mappings to the factors, they still obtain out-of-sample predictive R^2 and Sharpe ratio gains in relation to benchmark methods. Other notable uses of deep learning within factor model structures are Chen, Pelger, and Zhu (2020); Feng et al. (2018); Guijarro-Ordonez, Pelger, and Zanotti (2021); among others.

A fundamental difference between common empirical settings for machine learning applications and their use in empirical asset pricing is the uniquely low signal-to-noise DGP. The-

ory suggests market efficiency prices in the signal, such that, the unforecastable idiosyncratic error dominates. This critical issue significantly compounds the curse-of-dimensionality of using high-dimensional asset characteristics, which further motivates the parsimonious specification of the factor model.

1.3 Setup

Setting and Observable Random Variables Assume for time periods $t = 1, 2, \ldots, T$ and assets $i = 1, 2, \ldots, N$, that we observe realizations of random variables for asset excess returns $r_{i,t+1} \in \mathbb{R}$ and asset characteristics $z_{i,t} \in \mathbb{R}^p$. An asset's excess return is the simple return of asset i from time t to $t+1$ net the assumed simple return of the risk-free rate (e.g., one-month US Treasury Bill). An asset characteristic of asset i is known at time t : for example, the total fees for a crypto protocol between time $t-1$ and t. Note that asset characteristics are information from the previous period to follow the established convention in the literature and to be able to use this model for prediction. Importantly, we will introduce the novel asymptotic assumption for dynamic latent-factor models to let p grow to infinity simultaneously with N and T.

Model Given the highly nonlinear data-generating process observed in empirical asset returns (Gu, Kelly, and Xiu (2020); Chen, Pelger, and Zhu (2020); Bianchi, Büchner, and Tamoni (2021)), we specify a semi-parametric factor model—where $\beta_{i,t}$ is a function of asset characteristics $z_{i,t}$—studied in recent literature (Connor and Linton (2007); Connor, Hagmann, and Linton (2012); and Fan, Liao, and Wang (2016)). We assume a dynamic latent-factor model where

$$r_{i,t+1} = \beta_{i,t}^\top f_{t+1} + \epsilon_{i,t+1}^r.$$

However, given we are interested in conducting inference, we specify a linear model—to

build on the foundational work of Kelly, Pruitt, and Su (2019)—for the factor loadings

$$\beta_{i,t} = \Gamma_\beta^\top z_{i,t} + \epsilon_{i,t}^\beta.^{2}$$

Parameters and Unobserved Random Variables $f_{t+1} \in \mathbb{R}^k$ are low-dimensional latent factors; $\beta_{i,t} \in \mathbb{R}^k$ are latent-factor loadings; $\Gamma_\beta \in \mathbb{R}^{p\times k}$ is an unknown factor loading parameter matrix; and, $\epsilon_{i,t+1}^r \in \mathbb{R}$ and $\epsilon_{i,t}^\beta \in \mathbb{R}^k$ are unobserved idiosyncratic errors.

The latent factors f_{t+1} should be interpreted as purely statistical in nature. That is, these risk factors do not necessarily capture fundamental shocks to productive technologies as modeled in canonical theoretical models. Nevertheless, the latent factors capture systemic risk or covariance among asset returns that is non-diversifiable. We follow the literature in restricting k to be a small finite constant (i.e., $k \in \{1, 2, 3, 4, 5\}$) that, in our asymptotic theory, does not grow with p, T, N. It should be noted that, although ubiquitous, it is nevertheless a strong assumption: the empirical content of asset returns can be captured by a small number of strictly time-varying systematic risk factors.

The specification of $\beta_{i,t}$ provides several benefits. First, we enable the use of a dynamic loading to model a changing relationship (e.g., regime changes) between the cross section of returns and systematic risk. Yet, we reduce the parameter space from a $N \times T$ loading matrix β to the $p \times k$ loading matrix Γ_β. Second, we incorporate additional data from the large number of asset characteristics to influence the factor model for returns through the loading matrix Γ_β. This addresses a challenge of migrating assets wherein an asset-specific but static β_i would not be able to capture an asset moving from, for example, a crypto asset earning low fees to one with high protocol fees. The classic way to handle this issue was to sort assets into portfolios of similar characteristics to form test assets, which then compresses the dimensionality of the cross-section. Thus, as discussed in Kelly, Pruitt, and Su (2019),

[2] Although we are working with this parametric specification for the factor loadings, we can in our setting, nevertheless, employ feature engineering to generate many different functional forms of our asset characteristics, given the coming dimensionality reduction from LASSO.

this model specification skips ad hoc test asset formation to instead accommodate working directly with the high-dimensional system of individual assets. Finally, we assume exact row sparsity in Γ_β—precisely stated in the coming Assumption 1(ii)—a novel assumption to the literature; that is, only a small number of the p asset characteristics determine the content of the factor loading, which matches empirical findings in cross-sectional asset pricing (Kelly, Pruitt, and Su (2019); Babiak and Bianchi (2021); Bianchi, Guidolin, and Pedio (2022)).

Extended Model In order to show the generality of this approach, we enrich the model—with one of several possible extensions—to address the common question in asset pricing research of whether an observable factor $g_{t+1} \in \mathbb{R}$ carries a risk premium: compensation for exposure to the risk factor holding constant exposure to all other sources of risk, i.e., variation with other factors.

In the context of asset pricing, a factor can be either tradable or nontradable. A tradable factor is a portfolio, that is, a convex combination of tradable asset returns. The risk premium is straightforward to calculate for tradable factors: it is the time series average excess return of the factor. However, many risk factors are nontradable, e.g., inflation expectations, consumption, liquidity, etc. Thus, we must estimate the risk premia of nontradable observable risk factors as the risk premia associated with their tradable portfolio.

Following a recent approach in the literature (Giglio and Xiu (2021) and Giglio, Xiu, and Zhang (2021)), first, we assume the aforementioned model for returns is a function of the "true" latent factors f_{t+1}, [3] and, second, we assume these true latent factors can be decomposed into the latent-factor risk premia $\gamma \in \mathbb{R}^k$ (i.e., unknown parameters of the long-run average excess return) and latent-factor innovations $v_{t+1} \in \mathbb{R}^k$ (i.e., mean zero risk factor random variable), that is, $f_{t+1} \coloneqq \gamma + v_{t+1}$. Then, we specify the observable factor g_{t+1} as

[3] To be precise, by true latent factors, we mean we can consistently estimate the finite constant of the number of latent factors that span the cross section of returns.

potentially linearly correlated with the latent factors through

$$g_{t+1} = \eta v_{t+1} + \epsilon^g_{t+1},$$

where $\eta \in \mathbb{R}^k$ is an unknown parameter mapping the relation between the latent-factor innovations and the observable factors, and $\epsilon^g_{t+1} \in \mathbb{R}$ is measurement error in g_{t+1}. This specification allows the observable factors to be either simply some component(s) of the true latent factors (e.g., setting ϵ^g_{t+1} to zero with η set to $(1, 0, 0, \ldots, 0)$ or, more generally, some unknown linear function η of the true latent factors and thus still carry a risk premium. To recover the tradable portfolio representing the nontradable observable risk factor, we map g_{t+1} through η onto the column space of the true latent factors.

Precisely, the risk premium of an observable factor—our target parameter in this extension—is defined to be the expected excess return of a portfolio with loading (i.e., beta) of 1 with respect to this factor in g_{t+1} and zero loadings on all other factors; in this model, that parameter is $\gamma_g \coloneqq \eta^\top \gamma$.

Goal Under the novel asymptotic assumption for this setting of $p, N, T \to \infty$, we aim to develop estimation procedures for the latent loadings Γ_β and factors $f_{t+1}, \ \forall t$; in addition, we aim to estimate and conduct inference on γ_g under the novel use of a dynamic latent-factor model and the aforementioned novel high-dimensional asset characteristics.

1.4 Estimation

Motivating DSLFM Estimation The goal is to jointly estimate the loading matrix Γ_β and latent factors f_{t+1}, which are not separately identifiable without further restrictions, to be discussed (Bai and Ng 2013). However, to begin, given the model takes the form

$$r_{i,t+1} = z^\top_{i,t} \Gamma_\beta f_{t+1} + \epsilon_{i,t+1}$$

where $\epsilon_{i,t+1} = (\epsilon^{\beta}_{i,t})^{\top} f_{t+1} + \epsilon^{r}_{i,t+1}$ is the composite idiosyncratic error, we observe that our setting is high-dimensional panel data where we project $\{r_{i,t+1}\}_{i=1,t=1}^{i=N,t=T}$ onto the column space of $\{z_{i,t}\}_{i=1}^{N}$ for each time period to estimate each p dimensional time-varying vector $\{\widehat{\Gamma_{\beta} f_{t+1}}\}_{t=1}^{T}$ where we have to address $p \sim \max(N,T)$ or $p >> \max(N,T)$.

Thus, if we utilize the objective function to minimize over the $p-$dimensional choice vector $\Gamma_{\beta} f_{t+1}$ the mean-squared error of $\sum_i (r_{i,t+1} - z_{i,t}^{\top} \Gamma_{\beta} f_{t+1})^2$, we will not only have a noisily estimated design matrix when $p \sim \max(N,T)$, (or, at worst, a nonsingular design matrix when $p > \max(N,T)$) but also a non-convex objective given the interaction between minimization arguments f_{t+1} and Γ_{β}. This rules out implementing low-dimensional (in p) methods.

One potential solution would be to introduce sparsity in $z_{i,t}$, given that empirical estimates show, ex post, few covariates contribute the vast majority of the signal, as stated earlier.[4] This would amount to adding a regularization parameter to the aforementioned objective to combat the curse of dimensionality from $z_{i,t}$ [5] However, although potentially helpful for minimizing MSE by decreasing the variance of the estimator, this regularization introduces a bias in estimation, which will lead to invalid asymptotic inference for asset pricing tests, defeating a goal of this work.

We therefore adapt for our purpose the Double Selection Lasso estimator introduced by Belloni, Chernozhukov, and Hansen (2014). The key insight from their work was to introduce an orthogonality wherein, assuming Γ_{β} is row sparse, the regularization bias from the LASSO first-stage estimation does not pass through to the target parameter of interest when conducting inference.[6] First, we estimate for each time period t and each characteristic

[4]The online implementation of IPCA Kelly, Pruitt, and Su (2019) does indeed offer an ℓ_1 regularization to their MSE objective, which is not discussed in the econometric theory of Kelly, Pruitt, and Su (2020).

[5]One of several ways to interpret the curse of dimensionality is that as the number of covariates increases linearly, the volume of the parameter space to estimate grows nonlinearly; hence the density of the data falls.

[6]The ideas developed in the Double Selection Lasso paper for inference in partially linear models with high-dimensional controls were the basis for the generalization of this idea in the DML procedures as developed in

j the scalar $\Gamma_{\beta,j}^{\top} f_{t+1}$ using DSL; then stacking these estimates into a $T \times p$ matrix, we use PCA to obtain separate estimates for latent loadings $\widehat{\Gamma}_{\beta}$ and factors $\{\widehat{f}_{t+1}\}_{t=1}^{T}$; and, finally, soft-threshold $\widehat{\Gamma}_{\beta}$ to set the majority of the rows to zero given the assumption of sparsity.

We rewrite the DSLFM model and introduce a first-stage:

$$\begin{aligned} r_{i,t+1} &= z_{i,t,j} c_{t+1,j} + z_{i,t,-j}^{\top} c_{t+1,-j} + \epsilon_{i,t+1}, \qquad \mathbb{E}[\epsilon_{i,t+1} | z_{i,t}] = 0, \\ z_{i,t,j} &= z_{i,t,-j}^{\top} \delta_{t,j} + \epsilon_{i,t,j}^{z}, \qquad \mathbb{E}[\epsilon_{i,t,j}^{z} | z_{i,t,-j}] = 0, \end{aligned} \tag{1.2}$$

where $c_{t+1,j}$ refers to the $j \in \{1, \ldots, p\}$ component of $c_{t+1} := \Gamma_{\beta} f_{t+1}$ while $-j$ refers to the remaining $p-1$ elements of the vector; $\delta_{t,j} \in \mathbb{R}^{(p-1)}$ is an unknown, possibly time-varying, parameter; and, $\epsilon_{i,t,j}^{z}$ is an unknown scalar random error. $c_{t+1,j}$ is an asset return when its $j-$th characteristic is set to 1 and all other characteristics are set to zero, less its idiosyncratic return $\epsilon_{i,t+1}$.

There are several ways to interpret and justify the first-stage equation as discussed in Belloni, Chernozhukov, and Hansen (2014). Intuitively, the procedure does not rely on perfect model selection for valid inference as instead we not only recover controls $z_{i,t,-j}$ in the second-stage equation for their pricing ability in the cross section of returns but also recover controls with high correlation to our target variable $z_{i,t,j}$. From a theoretical perspective, the first-stage equation accounts for potential omitted variable bias if one estimated only the second-stage equation. That is, the set of potentially relevant asset covariates is enormous (Chen, Pelger, and Zhu (2020) and Bianchi, Guidolin, and Pedio (2022)), and thus a researcher may be motivated to select their preferred subset to ameliorate the curse of dimensionality, which could introduce model selection mistakes. Moreover, it is known LASSO can

Chernozhukov et al. (2018a). It would likely be closer to the empirical reality to maintain a nonparametric loading (Fan, Liao, and Wang 2016). As aforementioned, there is thus a natural extension of the work herein to use DML wherein the target variable $c_{t,j}$ is linear while the controls are nonparametrically estimated via a machine learning method, which would require the development of a Neyman Orthogonal score for this panel data setting, perhaps in a similar fashion to Semenova and Chernozhukov (2021a). However, the econometric theory is unknown for inference in the nonparametric setting. Moreover, given the high-dimension asset characteristics, using a conditional independence assumption to obtain a causal parameter may be the most fruitful path toward a causal factor model, a major area of future work (e.g., Lopez de Prado (2022)). We explore the out-of-sample predictive ability of a non-parametric model in the last section of this manuscript.

have poor finite sample model selection performance (Chernozhukov, Hansen, and Spindler 2015). Thus, selecting covariates with only the second-stage equation could fail to include relevant controls.

DSLFM Estimation Procedure Our estimation procedure for $\{f_{t+1}\}_{t=1}^{T}$ and Γ_β has three steps: Double Selection Lasso (DSL), PCA, and soft-threshold step.

1. *DSL*: To estimate $\hat{c}_{t+1,j}$, run $T \times p$ DSL cross-sectional regressions.[7]
   - Run LASSO of $\{r_{i,t+1}\}_{i=1}^{N}$ on $\{z_{i,t}\}_{i=1}^{N}$ for $\hat{c}_{t+1,j}$ and $\hat{c}_{t+1,-j}$.
     - Let $\hat{I}_1$ denote the nonzero elements of $\hat{c}_{t+1,-j}$.
   - Run LASSO $\{z_{i,t,j}\}_{i=1}^{N}$ on $\{z_{i,t,-j}\}_{i=1}^{N}$ for $\hat{\delta}_{t,j}$.
     - Let $\hat{I}_2$ denote the nonzero elements of $\hat{\delta}_{t,j}$.
   - Define the set $\hat{I} = \hat{I}_1 \cup \hat{I}_2 \cup \hat{I}_3$ where $\hat{I}_3$ is the set of controls in $z_{i,t,-j}$ not included in the first two LASSOs that the econometrician thinks are important for ensuring robustness, termed the amelioration set.
   - Run OLS of $\{r_{i,t+1}\}_{i=1}^{N}$ on $\{z_{i,t,j}, \tilde{z}_{i,t,-j}\}_{i=1}^{N}$ where $\tilde{z}_{i,t,-j}$ includes only elements of $z_{i,t,-j}$ in $\hat{I}$. That is,

$$(\hat{c}_{t+1,j}, \hat{c}_{t+1,-j}) \coloneqq \arg\min_{c_j, c_{-j}} \{\mathbb{E}_N[(r_{i,t+1} - z_{i,t,j}c_{t+1,j} - z_{i,t,-j}^\top c_{t+1,-j})] \ : \ c_{t+1,-j,l} = 0, \forall l \notin \hat{I}\}.$$

2. *PCA*: To estimate Γ_β and f_{t+1}, run PCA on $\hat{C} = \widehat{F}\widehat{\Gamma}_\beta^\top$—a $T \times p$ matrix—to decompose it into $p \times k$ and $T \times k$ matrices $\hat{\Gamma}_\beta$ and $\widehat{F}$.

3. *Soft-threshold*: Given the assumed exact row sparsity of Γ_β, we set to zero all rows of $\hat{\Gamma}_\beta$ whose row ℓ_1 norm is below a cross-validated hyperparameter λ. That is,

$$\check{\Gamma}_{\beta,j} = \left(\left\|\widehat{\Gamma}_{\beta,j}\right\|_1 - \lambda\right)_+ \text{sign}(\left\|\widehat{\Gamma}_{\beta,j}\right\|_1), \quad j \in \{1, \ldots, p\}. \tag{1.3}$$

[7] To set the penalty parameter in the LASSO implementations, one can follow the analytic method developed for heteroskedastic, non-Gaussian settings detailed in Appendix A, Algorithm 1 of Belloni, Chernozhukov, and Hansen (2014). For a more modern approach, one can use the bootstrap-after-cross-validation method of Chetverikov and Sorensen (2021). In practice, we use cross validation.

This does require running $T \times p$ versions of the cross-sectional Double Selection Lasso regressions, which can be in the thousands in empirical settings; however, these regressions are all computationally light and can be run in parallel. Moreover, this allows for unbalanced panels as each cross-section can have a different number of assets. [8] Additionally, these cross-sectional regressions, followed by estimations with the entire panel, mirror the effort of the most commonly used estimation procedure in the factor model setting, namely two-pass Fama-MacBeth regressions.

The high dimensionality of the PCA procedure, given we have a $p \times T$ matrix, is adapted from the existing literature using $N \times T$ matrices (Bai 2003). The estimated factor matrix $\widehat{F}$ is the product of $\sqrt{T}$ and the eigenvectors corresponding to the k largest eigenvalues of the $T \times T$ matrix $(Tp)^{-1}\widehat{C}\widehat{C}^\top$. The estimated factors are normalized such that $\widehat{F}^\top\widehat{F} = I_{k\times k}$, a standard approach. The estimated loadings are $\widehat{\Gamma}_\beta = T^{-1}\widehat{C}^\top\widehat{F}$. We thus see the main challenge in deriving the large-sample theory will be handling the estimation error in using $\widehat{C}$ instead of the unobserved C.

The final soft-thresholding step 1.3 in our estimation procedure exploits the sparsity in Γ_β to not only reduce the dimensionality of the characteristic space s ($<< p$) but also remove noise from the characteristics that have low signal-to-noise ratios. Belloni et al. (2018) discuss the general theoretical properties of the soft-threshold estimator with theory-based hyperparameter selection, and its close relation, the better known LASSO and Dantzig selector estimators. We further discuss constraints and selection of the hyperparameter in Appendix 1.8.1.

Estimating the Risk Premium of an Observable Factor Under the richer setting that includes the observable factor g_{t+1}, our model has an additional specification and moment

[8] In empirical practice, we find the entire estimation procedure is on the order of ten minutes.

conditions.

$$\begin{aligned} r_{i,t+1} &= z_{i,t}^\top \Gamma_\beta(\gamma + v_{t+1}) + \epsilon_{i,t+1}, \quad \mathbb{E}[v_{t+1}] = \mathbb{E}[\epsilon_{i,t+1}] = 0, \quad \mathbb{E}[v_{t+1}\epsilon_{i,t+1}] = 0, \\ g_{t+1} &= \eta v_{t+1} + \epsilon^g_{t+1}, \qquad\qquad\qquad\qquad \mathbb{E}[\epsilon^g_{t+1}] = 0, \quad \mathbb{E}[v_{t+1}\epsilon^g_{t+1}] = 0. \end{aligned} \tag{1.4}$$

Our goal is to estimate and conduct inference on $\gamma_g \coloneqq \eta^\top \gamma$. Given the latent factors are unobserved, we cannot jointly estimate v_{t+1} and η without further restrictions. We would have to invoke one of the three classic identification approaches of Bai and Ng (2013); however, by using the key rotation invariance result of Giglio and Xiu (2021), we can estimate the latent factors up to an invertible rotation matrix $H \in \mathbb{R}^{k\times k}$ and still maintain identification of our target parameter γ_g. That is, both of the underlying parameters will be identified up to this rotation matrix: $\eta^\top = \eta_0^\top H^{-1}$ and $\gamma = H\gamma_0$. Thus, the target parameter is rotationally invariant to H : $\gamma_g = \eta_0^\top H^{-1} H \gamma_0 = \eta^\top \gamma$.

For our estimation procedure, we replace the first PCA step of Giglio and Xiu (2021) with our above procedure—augmented to use return innovations—to estimate the latent loadings $\check{\Gamma}_\beta$ and factor innovations $\hat{v}_{t+1}$ for all t. We then proceed with the latter two steps of the authors' procedure to obtain our target estimator.

1. To estimate latent-factor risk premia $\hat{\gamma}$, run cross-sectional OLS of average returns $\bar{r} \in \mathbb{R}^N$ on averaged estimated latent-factor loadings $\bar{\hat{\beta}} = \bar{Z}^\top \hat{\Gamma}_\beta \in \mathbb{R}^{N\times k}$.

2. To estimate latent to observable factor mapping $\hat{\eta}$, run a time series OLS regression of $\{g_{t+1}\}_{t=1}^T$ on factor innovations $\hat{V} \in \mathbb{R}^{T\times k}$.

We can thus form our estimator of the risk premium for the observable factors g_{t+1} by combining these estimators into $\hat{\gamma}_g = \hat{\eta}^\top \hat{\gamma}$.

This procedure extends the estimation in Giglio, Xiu, and Zhang (2021) to dynamic loadings and high-dimensional asset characteristics, while inheriting the rotation invariance and the specification consistent with two-pass estimators in this literature. Again, simply applying IPCA instead of PCA in the first step of Giglio and Xiu (2021) would not be feasible

with $p > \max\{N, T\}$ or would yield biased inference if an ℓ_1 penalty were simply added to the IPCA objective. The cross-sectional OLS of average returns on the estimated latent-factor loadings is the standard second step in two-pass Fama-MacBeth regressions, which could be replaced with generalized least squares or weighted least squares to explore asymptotic efficiency gains. The final time series regression is critical to translate the uninterpretable risk premia of latent factors to those of factors proposed by economic theory. Moreover, this procedure handles omitted variable bias which we now briefly discuss.

To illustrate, assume we have a scalar observable factor g_{t+1}, which is the first component of a two-dimensional latent-factor innovation vector: $v_{t+1} = (g_{t+1}, v_{2,t+1})^\top$ (i.e. $\eta = (1, 0)$). The vector-version of our model is thus

$$r_{t+1} = z_t \Gamma_{\beta,1}(\gamma_g + g_{t+1}) + z_t \Gamma_{\beta,2}(\gamma_2 + v_{2,t+1}) + \epsilon_{t+1}.$$

Using the standard Fama-MacBeth two-pass regressions (Fama and MacBeth 1973) will produce bias in estimating γ_g if $v_{2,t+1}$ is omitted. The first step of a time series regression of asset excess returns on g_{t+1} will give a biased estimate of $\hat{\beta}_1$ as long as $v_{2,t+1}$ is correlated with both g_{t+1} and r_{t+1}, per the standard OVB term: the covariance between between the outcome and the excluded regressor times the covariance between the included and excluded regressor, up to scale. Moreover, in the second step of a cross-sectional regression of average returns on estimated loadings, a second omitted variable bias is introduced if the loading of the omitted factor $\hat{\beta}_2$ is correlated with both $\hat{\beta}_1$ and $\bar{r}_{t+1}$.

Estimating the latent factors via the DSLFM procedure resolves this issue of omitting a potentially relevant factor given one can utilize a consistent estimator of the true number of latent factors, which we assume spans the true factor space.[9]

[9]The DSLFM could be further extended to estimate the zero-beta rate (i.e., alpha) using a very similar approach to that discussed in Online appendix I.2 of Giglio and Xiu (2021).

1.5 Asymptotic Theory

In this section, we present the asymptotic results for consistent estimation of the latent factors and loadings and the large sample distribution of the nontradable observable factor risk premium estimator under the assumed setting discussed in Section 1.3 and using estimation procedures discussed in Section 1.4 for models (1.2) and (1.4). We first provide the regularity conditions sufficient for the validity of the estimation and inference results. For clarity of exposition, we focus on motivating the assumptions and interpreting the results, while theoretical details and mathematical proofs are provided in Appendix 1.8.1.

Throughout, let $\|A\| = [tr(A^\top A)]^{1/2}$ denote the Frobenius norm of matrix A or $\|x\| = (\sum_i x_i^2)^{1/2}$ for the ℓ_2 norm of a vector x. Let $\|x\|_0$ and $\|x\|_1$ be the usual ℓ_0 and ℓ_1 norms, respectively. All limits are simultaneous where we will restrict the rates among p, T, N, to allow $p \to \infty$, as discussed below.

1.5.1 Regularity Conditions

Consistent Estimators for the Latent-Factor Model The following assumptions enable the consistent estimation of the factors $\{f_{t+1}\}_{t=1}^T$ and the loadings Γ_β. Let f_{t+1}^0 and Γ_β^0 be the true factors and loadings such that $f_{t+1} = Hf_{t+1}^0$ and $\Gamma_\beta = \Gamma_\beta^0 H^{-1}$ where H is an unobserved $k \times k$ invertible rotation matrix.

In regard to identification, our results do not require the identification of the true factors f_{t+1}^0 and loadings Γ_β^0 but rather simply factors (loadings) that span the true factors (loadings) up to the rotation matrix H. Bai and Ng (2013) show identification results for PCA under three different sets of assumptions to pin down the $k \times k$ elements in H, which requires pinning down the covariance matrices of the factor loadings and factors to be diagonal matrices or identity matrices to provide $k(k-1)/2 + k(k+1)/2 = k^2$ restrictions. The researcher can choose which asymptotic covariance matrix to restrict. As we will discuss, we will additionally not need these identification restrictions for the observable factor risk

premia given the aforementioned rotation invariance result of the target parameter.

ASSUMPTION 1 (Consistency of DSL). 1. *Bounded Characteristic Portfolios: For a finite absolute constant* M *and* $\forall t, j$, $|c_{t+1,j}| = \left|\Gamma_{\beta,j}^\top f_{t+1}\right| < M$.

2. *Sparse Loading: Loading matrix* Γ_β *admits an exactly sparse form. That is, for* $\exists s \in \mathbb{N}_+, i.e. p > s \geq 1$, Γ_β *has at most* s *nonzero rows:* $\sum_{j=1}^p \mathbb{1}\Big\{\|\Gamma_{\beta,j}\|_1 > 0\Big\} \leq s$.

These are two critical assumptions for DSL consistency with the additional standard and technical DSL assumptions in Appendix 1.8.1. Assumption 1(i) converts the bounded target parameter, in the traditional DSL context, to the DSLFM context where we require realizations of $c_{t+1,j}$ to be finite-sample bounded by a constant that does not depend on p, T, N. This imposes a bound on the return of characteristic portfolios, that is, the return of a portfolio with characteristic j set to 1 and all other characteristics set to 0. We could instead assume returns are bounded random variable to impose Assumption 1(i).

Assumption 1(ii) is the key LASSO assumption that the parameter on the control regressors admits an exactly sparse form, which follows from our assumption such that $\forall t, j$, $\|c_{t+1,-j}\|_0 = \|\Gamma_{\beta,-j} f_{t+1}\|_0 \leq s$. This sparsity of the loading matrix is supported empirically in asset pricing given the relevance of only a small number of asset characteristics, which we corroborate in our empirical setting. We have thus adapted the classic LASSO sparsity assumption to the empirical reality of cross-sectional asset pricing using high-dimensional asset characteristics. Exact sparsity could be relaxed to approximate sparsity with a similar but alternative high-dimensional econometrics toolkit.

We next turn to assumptions for consistently estimating the latent factors and loadings. The focus in our work is controlling the estimation error between the infeasible eigendecomposition of $(Tp)^{-1}CC^\top$ and the feasible eigendecomposition of $(Tp)^{-1}\widehat{C}\widehat{C}^\top$, given we do not observe $C = F\Gamma_\beta^\top$ but instead estimate each element via DSL and then eigendecompose using standard PCA estimators as discussed in Section 1.4.

Assumption 2 (Consistency of Latent-Factor Model). *1. Factors:* $\mathbb{E}\left\|f_{t+1}^0\right\|^4 \leq M < \infty$ *and* $T^{-1}\sum_t f_{t+1}^0 f_{t+1}^{0\top} \to_p \Sigma_f$ *for some* $k \times k$ *positive definite matrix* Σ_f.

2. *Factor Loadings:* $\forall j,\ \|\Gamma_{\beta,j}\| \leq M < \infty$ *and* $\left\|\Gamma_\beta^\top \Gamma_\beta / p - \Sigma_\Gamma\right\| \to 0$ *for some* $k \times k$ *positive definite matrix* Σ_Γ.

3. *Nonzero and distinct eigenvalues: from the infeasible eigendecomposition, the* k *largest eigenvalues* λ_i *for* $i \in \{1, \ldots, k\}$ *are bounded away from zero. Moreover, the* k *largest infeasible eigenvalues are distinct, that is,*

$$\min_{i:i\neq\kappa} |\lambda_\kappa - \lambda_i| > 0.$$

Assumptions 2(i)-(ii) are standard for factor models where the literature is styled after Assumptions A, B, and C of Bai (2003). Assumption 2(i) does not impose i.i.d. factors, as in the classical factor analysis literature, but instead imposes the factors are stationary, strong mixing, and satisfy moment conditions. Assumption 2(ii) ensures each latent factor contributes to the second moment of c_{t+1}; that is, it imposes all factors are pervasive and excludes weak factors. See Giglio, Xiu, and Zhang (2021) for adjustments for weak factors. The PCA estimation herein does not require the Assumption C of Bai (2003) given our target matrix $C = F\Gamma_\beta^\top$ is without an error term; we instead are controlling cross-sectional and temporal dependence using the moment conditions of DSL given in model (1.2) and more technical assumptions in Appendix 1.8.1.

Assumption 2(iii) assumes the $k-$largest eigenvalues from the infeasible and feasible eigendecompositions remain nonzero asymptotically. In finite sample, these are real and nonzero eigenvalues given we are taking the eigendecomposition of a rank k symmetric matrix. It is reasonable to assume we have distinct eigenvalues given, for this not to hold, there would have to be two or more dimensions in the $k-$largest of the $T \times T$ matrix $CC^\top$ that have precisely the same variability.

Estimating Number of Factors Given the focus of this work is on the consistency of the main estimators and the asymptotic distribution of the risk premium estimator, we assume $k = k^0$ is known.[10]

ASSUMPTION 3 (Consistent Estimator for Number of True Factors). *For $\bar{k} > k^0$, let $\widehat{k} \coloneqq \arg\min_{0 \leq k \leq \bar{k}} IC(k)$ where*

$$IC(k) \coloneqq \log(V(k)) + k\left(\frac{p+T}{pT}\right)\log\left(\frac{pT}{p+T}\right)$$
$$V(k) \coloneqq \min_{\Gamma_\beta, F} (pT)^{-1} \sum_{j,T} \left(c_{j,t+t} - \Gamma_{\beta,j}^\top f_{t+1}\right)^2.$$

Assume $\widehat{k} \to_p k^0$ without further restriction on the growth rates among p, T, N and $k = k^0$ is known.

Assumptions 1 and 2 can be shown to be sufficient for consistently estimating, with the above Information Criterion, the number of true factors k^0 using the results of Appendix 1.8.1 as in Bai and Ng (2002) and Bai (2003). Although providing this assumption to show the estimator to be studied in simulation, we are instead choosing to impose Assumption 3 in the asymptotic to focus on the main results of this work. Note that commonly used model selection criteria (e.g., AIC or BIC) will not yield consistent estimators, hence the specification above using the contribution of Bai and Ng (2002).

Inference on Nontradable Observable Factor Risk Premia The final assumptions are needed to derive the limiting distribution of the risk premium estimator.

[10]The asymptotic distribution of the risk premium estimator is unaffected when the number of factors is estimated because

$$\begin{aligned}\Pr\left(\widehat{\gamma}_g \leq x\right) &= \Pr\left(\widehat{\gamma}_g \leq x, \widehat{k} = k^0\right) + \Pr\left(\widehat{\gamma}_g \leq x, \widehat{k} \neq k^0\right) = \Pr\left(\widehat{\gamma}_g \leq x, \widehat{k} = k^0\right) + o(1) \\ &= \Pr\left(\widehat{\gamma}_g \leq x | \widehat{k} = k^0\right)\Pr\left(\widehat{k} = k^0\right) + o(1) = \Pr\left(\widehat{\gamma}_g \leq x | \widehat{k} = k^0\right) + o(1).\end{aligned}$$

ASSUMPTION 4 (Inference). *There exists a generic absolute constant $M < \infty$ such that for all p, T, N :*

1. *Bounded idiosyncratic errors:* $\mathbb{E}[(\sum_t \epsilon_{i,t+1})^2] \leq TM$.
2. *Bounded scaled factor innovations:* $\mathbb{E}[(\sum_t z_{i,t}^\top \Gamma_\beta^0 v_{t+1}^0)^2] \leq sTM$.
3. *Bounded measurement errors:* $\mathbb{E}[(\epsilon_{t+1}^g)^2] \leq M$.
4. *Convergence of characteristics:* $\frac{1}{NT}\sum_i \sum_{t'} \mathbb{E}[z_{i,t,j}] z_{i,t',j'} \to_p \mathcal{Z}_{t,j,j'}$ *uniformly over* t, j, j' *for* $j, j' \in \{1, 2, \ldots, p\}$ *and a nonstochastic finite constant* $\mathcal{Z}_{t,j,j'} \in \mathbb{R}$.
5. *CLT: As* $T \to \infty$, *the following joint central limit theorem holds:*

$$\frac{\sqrt{T}}{T} \sum_t \begin{pmatrix} v_{t+1}^0 \epsilon_{t+1}^g \\ \Pi_t v_{t+1}^0 \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Phi)$$

where random matrix $\Pi_t \in \mathbb{R}^{k \times k}$ *and nonstochastic matrix* $\Phi \in \mathbb{R}^{2k \times 2k}$ *are defined in Appendix 1.8.1.*

Assumption 4(i) bounds the second contemporaneous and cross-moments of the idiosyncratic errors, aligning with the time and cross-section dependence assumptions of Bai (2003) Assumption C. The assumption would hold if we assumed $\epsilon_{i,t+1}$ are uncorrelated across t, which is a simplified yet plausible assumption given the low signal-to-noise environment of asset pricing. We have thus relaxed the temporal dependence to the specified rate T.

Assumption 4(ii) bounds the squared time series average of the factor innovations scaled by the factor loadings. In the static factor model context of Giglio and Xiu (2021), this holds in large sample by a simple LLN argument given the static loadings are not a function of t and the factor innovations are mean zero random variables. Thus, we are ensuring the Γ_β^0 selected columns of Z_t keep the scaled v_{t+1}^0 sufficiently small.

Assumption 4(iii) bounds the second moment of the observable factor measurement errors for use in proving $\|\epsilon^g\| = O_p(\sqrt{T})$. It is not a stringent assumption because we are bounding

a zero mean scalar random variable. This is nearly an identical assumption and usage to Giglio and Xiu (2021) Assumption A8.

Assumption 4(iv) provides a convergence result such that the squared first moment for two different characteristics averaged over time and across assets is a nonstochastic finite constant. This is a weaker assumption on the distribution of characteristics than the DSL moment conditions discussed in Appendix 1.8.1.

Assumption 4(v) is the assumed central limit theorem for the $2k$ (low) dimensional mean zero random variable given the models' 1.2 and 1.4 moment assumptions, which is satisfied by various mixing processes. The second moments of the later $2k$ random variables are bound already in Assumptions 4(i)-(ii). We nevertheless directly assume the needed CLT. This extends for our inference result the assumed CLT at the same rate in Assumption F4 of Bai (2003) and the assumed CLT at the same rate in Assumption A11 of Giglio and Xiu (2021). Note that although we have the same two mean zero random vectors, our factor innovations are scaled by Π_t instead of a constant 1 given the dynamic factor loadings of our model.

1.5.2 Theory Results

This section presents the three main theoretical results.

Consistent Estimators for the Latent-Factor Model We present the first two results showing the consistency of the latent-factor model estimators.

Proposition 1 (Consistency of Latent Factors). *Under the model (1.2), Assumptions 1, 2, 3, and DSL Assumptions in Appendix 1.8.1.2 where $T, N, p \to \infty$, then for all t the latent-factor estimator described above has the property that*

$$\widehat{f}_{t+1} - H^\top f^0_{t+1} = O_p\left(\sqrt{\frac{s\log(Tp)}{N}}\right).$$

The proof is in Appendix 1.8.1. This result establishes the convergence rate of the latent factor estimator in a dynamic latent-factor model with high-dimensional characteristics. If the factor loadings were static and known, β_i^0 for all i, then, f_{t+1}^0 would be estimated via a cross-sectional least squares with a convergence rate of $\sqrt{N}$. Bai (2003) establishes in Theorem 1(ii), under $N, T \to \infty$ for a static latent-factor model, the foundational result of a convergence rate of $\min\left(\sqrt{N}, T\right)$ for the consistency of the latent-factor estimator for the rotated true factors $H^\top f_{t+1}^0$. Incorporating dynamic loadings parameterized by high dimensional characteristics comes at the cost of slowing the rate to $\sqrt{pN/s\log(Tp)}$, which is nevertheless still reasonable for typical values of p, T, N. Additional standard DSL rates, which are less restrictive, are in Appendix 1.8.1.

Our rate is primarily driven by the $\sqrt{N/\log(Tp)}$ rate uniform consistency over t and j of the DSL estimation error $|\widehat{c}_{t+1,j} - c_{t+1}, j|$ as shown in Lemma A1 in Appendix 1.8.1. Given the model for $C = F^0\Gamma_\beta^{0\top}$ contains no error, the eigendecomposition of the unobserved C is exact for F^0H as shown in Lemma A6; and, thus, the estimation error from using $\widehat{C}$ instead of C drives this first main result. The assumed sparsity in Γ_β^0 does improve the rate with the p/s ratio.

It is worth reiterating that under our setting F^0 and Γ_β^0 are not separately identifiable, hence the $k \times k$ invertible matrix transformation H appears in each asymptotic result. Similarly, $\widehat{F}\widehat{\Gamma}_\beta^\top$ is an estimator of the identifiable, rotation invariant common component C, which is corroborated by simulation results. Moreover, in many cases knowing F^0H is equivalent to knowing F^0; for example, the regressor F^0 will give the same predicted values as using F^0H as a regressor given they have the same column space.

Proposition 2 (Consistency of Latent-Factor Loadings). *Under the model (1.2), Assumptions 1, 2, 3, and DSL Assumptions in Appendix 1.8.1.2 where $T, N, p \to \infty$, then the latent loading estimator described above has the property that*

$$\left(\check{\Gamma}_\beta - \Gamma_\beta^0 H^{-1}\right) = O_p\left(\sqrt{\frac{s\log(Tp)}{N}}\right).$$

The proof is in Appendix 1.8.1. This result establishes the convergence rate of the latent loading estimator in a dynamic latent-factor model with high-dimensional characteristics. When the factors, f^0_{t+1}, for all t, are observable, static loadings β^0_i can be estimated by a time series regression with a convergence rate of $\sqrt{T}$. Bai (2003) establishes in Theorem 2(ii), under $N, T \to \infty$ for a static latent-factor model, the foundational result of a convergence rate of $\min\left(N, \sqrt{T}\right)$ for the consistency of the latent-factor loading estimator $\widehat{\beta}_i$ for the rotated true factor loadings $H^{-1}\beta^0_i$. Incorporating dynamic loadings parameterized by high dimensional characteristics comes at the cost of slowing the rate to $\sqrt{N/s\log(Tp)}$, which is nevertheless still reasonable for typical values of p, T, N.

The rate follows similar reasoning to that of the latent-factor estimator in Proposition 1. However, here we are presenting the rate of the final soft-threshold estimator—derived using recent results in high dimensional econometrics Belloni et al. (2018)—wherein we use the uniform estimation error between the eigendecomposition of $\widehat{C}$ for $\widehat{\beta}_{\beta,j}$ and the infeasible loading $\widetilde{\beta}_{\beta,j}$ from decomposing the unobserved C, which eliminates p in our rate from Proposition 1. The $\sqrt{N/log(Tp)}$ is similarly driven by the uniform consistency over t and j of the DSL estimation error $|\widehat{c}_{t+1,j} - c_{t+1}, j|$, which is the key result used to establish these consistency propositions along with typical high dimensional random matrix theory (e.g., Davis Kahan Theorem, Weyl Inequality, and recent tools in high dimensional econometric theory found in Belloni et al. (2018)).

Inference on Nontradable Observable Factor Risk Premium Finally, we present the asymptotic normality of the nontradable observable factor risk premium estimator.

THEOREM 1 (Normality of Observable Factor Risk Premium). *Under the models (1.2) and (1.4); Assumptions 1, 2, 3, 4; DSL Assumptions in Appendix 1.8.1.2; and if* $Ts^2\log(Tp)/N \to 0$, *then as* $T, N, p \to \infty$ *the estimator* $\hat{\gamma}_g$ *obeys*

$$\sqrt{T}\frac{(\hat{\gamma}_g - \gamma_g)}{\sigma_g} \xrightarrow{d} \mathcal{N}(0,1),$$

where σ_g *is defined in Appendix 1.8.1.*

The proof is in Appendix 1.8.1. This result establishes $\sqrt{T}$ asymptotic normality of the nontradable observable factor risk premium estimator from Giglio and Xiu (2021) extended to the setting of dynamic factor loadings with high dimensionality characteristics. At a high level, our proof follows a similar approach yielding, as seen in the proof of Theorem 1, the same two asymptotically nonnegligible terms as in Giglio and Xiu (2021). The first term arises from the time-series regression of the observed factor on the latent loadings where again the latent loading estimation error is higher order. The latter term in the $2k$ random vector in Assumption 4(v) arises from the cross-sectional regression of averaged asset excess returns on averaged factor loadings, where the factor loading estimation error and idiosyncratic error term are higher order. Although Giglio and Xiu (2021) have this same second term, ours is more complicated given the dynamic loadings, which necessitates the convergence Assumption 4(iv). A direct application of the delta method on the sum of these two terms yields the result in Theorem 1.

The crucial rate assumption is $Ts^2 \log(Tp)/N \to 0$, which controls the estimation error for the unobserved averaged latent-factor loadings $T^{-1}\sum_t \beta_{i,t}$. This is similar to Bai (2003) and Giglio and Xiu (2021), which require $T/N \to 0$ to use the estimated factors or loadings as generated regressors. However, we have slowed the rate again due to the high-dimensionality in p. This is our slowest required rate.

Given the rotation invariance of the target parameter γ_g^0, the unobserved rotation matrix H does not appear in the asymptotic distribution, in contrast to the consistency results. In finite sample, we corroborate this result in the to-be-discussed simulations. In a similar vein, it warrants noting the asymptotic efficiency loss due to not observing the factors and loadings could be large when N is relatively small. Also, we have assumed we directly observe the number of tree factors k^0, which would require estimation in practice and thus likely contribute estimation error to affect finite sample performance.

In simulation, we use the plug-in estimator for σ_g, which has satisfactory finite-sample coverage properties. However, one can establish a consistent variance estimator using a

Newey and West (1987) style plug-in estimator of the asymptotic variance σ_g with lag corrections to account for temporal dependence as in Giglio and Xiu (2021) Section IV Part E.

1.6 Asset Pricing Tests

In this section we develop three tests central to our empirical analysis. The first uses the asymptotic normality of the observable factor risk premium for a statistical test of nonzero risk compensation. The second statistic informs the incremental significance of any specific asset characteristic. The third and final discusses how we empirically measure whether the DSLFM contributes predictive signal above and beyond a random walk.

Testing Nontradable Observable Factor Risk Premium An empirical application of the DSLFM model will address whether a nontradable observable factor, namely, inflation, carries a nonzero risk premium in the crypto asset class. The target parameter γ_g captures the risk premium of the (inflation) factor-mimicking portfolio within the crypto asset class as recovered by the estimated dynamic latent-factor model. We are interested not only in the sign of the parameter, but also, in practical settings, in whether a confidence interval suggests a risk premium of economic significance.

We test the hypothesis $H_0 : \gamma_g = 0 \quad \text{vs.} \quad H_1 : \gamma_g \neq 0$ using the risk premium estimation procedure described in Section 1.4 with a plug-in variance estimator $\widehat{\sigma}_g$ for σ_g. Given the asymptotic normality of Theorem 1, we form a confidence interval

$$\gamma_g \in [\widehat{\gamma}_g - c(1-\alpha/2)\widehat{\sigma}_g, \widehat{\gamma}_g + c(1-\alpha/2)\widehat{\sigma}_g]$$

where the critical value $c(1-\alpha/2)$ is the $1-\alpha/2$ quantile of a $N(0,1)$ distribution for the researcher-specified level of the test α. We find in the coming simulation acceptable finite sample coverage for this confidence interval.

Testing Characteristic Significance The large-sample distribution of the latent loading is unknown given the DSL regularization. Even inference in simple cross-sectional LASSO is complicated (Lee et al. 2016). Instead, we develop a simple bootstrap procedure to infer whether a specific characteristic significantly contributes to loading Γ_β. We leave for subsequent research developing the supporting theory of this bootstrap procedure or develop the asymptotic distribution of a consistent latent loading estimator in this setting. We test the hypotheses $\forall j$

$$H_0 : \Gamma_\beta^\top = [\Gamma_{\beta,1}, \ldots, \Gamma_{\beta,j-1}, 0, \Gamma_{\beta,j+1}, \ldots, \Gamma_{\beta,p}] \quad \text{vs.} \quad H_1 : \Gamma_\beta^\top = [\Gamma_{\beta,1}, \ldots, \Gamma_{\beta,p}] .$$

That is, we ask whether characteristic j contributes to the factor loading through $k \times 1$ mapping vector $\Gamma_{\beta,j}$. This allows the researcher, using a large number of characteristics, to systematically ask what characteristics contribute to the latent-factor model, instead of an ad hoc selection. We thus set the entire $k \times 1$ vector to zero so the characteristic contributes to predicting the variation in returns through none of the k factors.

Our procedure is to test the alternative hypothesis model, with the unconstrained characteristic j, and then form the test statistic

$$W_{\Gamma,j} = \Gamma_{\beta,j}^\top \Gamma_{\beta,j}.$$

Using bootstrapped standard errors, we assess whether this test $W_{\Gamma,j}$ statistic is statistically distinguishable from zero.

Testing Out-of-Sample Performance To study the out-of-sample pricing ability of the DSLFM, we use the "predictive R^2" defined as

$$\text{Predictive } R^2 = \frac{\sum_{i,t} \left(r_{i,t+1} - z_{i,t}^\top \check{\Gamma}_\beta \widehat{\lambda}_t \right)^2}{\sum_{i,t} r_{i,t+1}^2}$$

where $\widehat{\lambda}$ is the moving average of the estimated factors in previous time periods over a cross-validated window size. This measure captures whether the model forecasts realized

returns better than a random walk; or, said differently, it represents the fraction of realized return variation explained by the model's description of expected returns through exposure to systematic risk. This specification allows the model's estimated conditional expected returns to be driven not just by the dynamic factor loadings, estimated using high-dimensional asset characteristics, but also by time-varying risk prices λ_t.

1.7 Simulations

This section presents a brief study of the finite-sample performance of the dynamic latent-factor model estimators and coverage properties of the inference procedure using Monte Carlo simulations. To summarize, we find the estimation errors for factors and loadings are comparable to IPCA and the Three Pass estimator of Giglio and Xiu (2021) in low-dimensional settings while superior in high-dimensional settings. This holds even in rather small samples with low signal to noise ratios, reflecting the empirical reality of cross-sectional asset pricing. Moreover, we find estimation errors and coverage properties for the observable-factor risk premium to be comparable to Giglio and Xiu (2021) in low-dimensional settings while superior in high-dimensional settings. We now present the design, followed by the results.

Simulation Design First, we describe the data-generating process for given N, T, k where we follow the finite sample simulation study of IPCA (Kelly, Pruitt, and Su 2020). That is, the DGP is favorable to IPCA. We calibrated the simulated data to parameter estimates from IPCA fit to our weekly panel of crypto asset data using all sixty three asset characteristics.

Latent factors f_{t+1} are simulated from a $VAR(1)$ model employing normal innovations that was fit to the estimated IPCA factors. Asset characteristics are simulated from a p variable panel $VAR(1)$ model with normal innovations, which was fit to the demeaned empirical weekly panel of randomly selected, without replacement, p asset characteristics.

For each asset, we set the means of the characteristics to a bootstrap sample from the empirical distribution of time series asset characteristic means. The idiosyncratic error $\epsilon_{i,t+1}$ is simulated from an i.i.d. normal distribution whose variance is calibrated such that the population R^2 of the model is approximately 20%, matching the empirically estimated value from fitting IPCA. The measurement error ϵ^g_{t+1} is simulated in a simple fashion but the $R^2 = 1 - \mathbb{E}[e^g]/\mathbb{E}[g]$ is calibrated to approximately 40%. $\eta = (1, 0, \ldots, 0)$ and the loadings Γ_β are set to the empirically estimated values where $p - s$ rows are set to zero where $s = p/10$. Finally, observable factors and returns are generated according to models (1.2) and (1.4).

The simulation studies results across $S = 200$ Monte Carlo draws. Hyperparameters are fixed at $N = 500$, $T = 100$, and $k = 3$. To compare the performance of estimators under low-dimensional and high-dimensional characteristics, results are generated for $p = 10$ and $p = 50$. We report results for a variety of estimators, including latent loadings Γ_β, latent factors F, average factor loadings $\bar{\beta}$, latent matrix $C = F\Gamma_\beta^\top$, and observable factor risk premium γ_g.

The benchmark estimation and inference procedures are IPCA and the three-pass estimator of Giglio and Xiu (2021), given DSLFM's basis on these foundational models.[11] We focus on two comparisons: first, the estimation error of theoretically consistent latent loading Γ_β and latent factor $\{f_{t+1}\}_{t=1}^T$ IPCA and DSLFM estimators; and, second, coverage properties of the observable factor risk premium estimator. We do study estimation errors for additional estimands as relevant (e.g., the three-pass estimator does not estimate latent loadings Γ_β nor latent matrix C, while IPCA does not have an observable factor risk premium estimation nor inference procedure).

Simulation Results Table 1.1 reports results. In the low-dimensional setting of $N = 500, T = 100, p = 10, s = 1$, we find DSLFM to obtain smaller estimation errors for Γ_β as

[11] Many thanks to Matthias Buechner and Leland Bybee for the IPCA implementation https://github.com/bkelly-lab/ipca.

compared to IPCA; however, IPCA has an order of magnitude lower estimation errors for F. DSLFM's outperformance for the latent loading is driven by taking on higher bias yet substantially lower variance of the estimator; this is obtained from soft-thresholding many of the rows. In comparing other auxiliary estimands, DSLFM obtains lower estimation error for the time-series averaged factor loadings $\bar{\beta}$ although higher error for the latent matrix $C = F\Gamma_\beta^\top$. We attribute these results to IPCA fitting data simulated to match the fits of an empirically estimated IPCA model, yet we employ an exact row-sparsity structure in the true latent loadings Γ_β.

DSLFM slightly under-covers the 90% and 95% confidence intervals in the low-dimensional setting. The three-pass estimator obtains similar estimation error for the target parameter of the observable factor risk premium, but has finite-sample intervals that slightly over-cover.

Moving to the high-dimensional setting of $p = 50$, we find DSLFM to again obtain smaller estimation errors for latent loadings, yet now the estimation errors for the latent factors are of the same order as IPCA. In both cases, DSLFM takes on bias from its regularization methods, although DSLFM's latent factor estimator's variance is still higher than IPCA. DSLFM is now an order of magnitude improvement for the average factor loadings and a factor of two improvement for the latent matrix C. Finally, coverage proprieties of the risk premium estimand for both the three-pass estimator and DSLFM estimators are degraded under high-dimensionality.

We hope to add even higher dimensional results with hyperparameters closer to the empirical values. Do note, given the DSLFM's large sample theory, it is constrained by $N > T$, which although is the case for the panel of crypto asset returns, this is not the case for all cross-sectional asset pricing settings. Moreover, the DSLFM's performance was boosted by the assumed exact sparsity in the latent loadings; we hope to add results for approximate sparsity, which is likely closer to the empirical reality. Nevertheless, the DSLFM performs well as compared to state-of-the-art benchmark methods, especially under the setting for which it was developed: high-dimensional asset characteristics.

1.8 Appendix

1.8.1 Technical Details and Proofs

1.8.1.1 Notation

Let $\mathbb{E}_N[x_i] \coloneqq N^{-1}\sum_t x_i$ for random variables $\{x_i\}_{i=1}^N$.

Let $\mathbb{I}_k$ be a $k \times k$ identify matrix. Let $\|\cdot\|$ be the Frobenius norm for a matrix and the ℓ_2 norm for a vector; $\|\cdot\|_1$ be the l_1-norm; $\|\cdot\|_2$ be the spectral norm for a matrix; and, $\|\cdot\|_\infty$ be the maximum element of the matrix or vector. Let $a \vee b = max(a, b)$. We also use the notation $a \lesssim_P b$ to denote $a = O_p(b)$ for $a, b \in \mathbb{R}$.

Define the following random variables: $r_{t+1} = (r_{1,t+1}, \ldots, r_{N,t+1})^\top \in \mathbb{R}^N$;
$z_{t,j} = (z_{1,t,j}, \ldots, z_{N,t,j})^\top \in \mathbb{R}^N$; $Z_{t,-j} = (z_{1,t,-j}, \ldots, z_{N,t,-j})^\top \in \mathbb{R}^{N\times(p-1)}$;
$\epsilon_{t+1} = (\epsilon_{1,t+1}, \ldots, \epsilon_{N,t+1})^\top \in \mathbb{R}^N$; $\epsilon^z_{t,j} = (\epsilon^z_{1,t,j}, \ldots, \epsilon^z_{N,t,j})^\top \in \mathbb{R}^N$, and so on.

For $A \subset \{1, \ldots, p\}$, let $Z_{t,-j}[A]$ denote the subset of the columns of $Z_{t,-j}$ that are elements of the set A. Let $\mathcal{P}_A \coloneqq Z_{t,-j}[A]\left(Z_{t,-j}[A]^\top Z_{t,-j}[A]\right)^{-1} Z_{t,-j}[A]^\top$ be the projection operator that maps vectors in $\mathbb{R}^N$ into $\text{span}(Z_{t,-j}[A])$. Let $\mathcal{M}_A = \mathbb{I}_N - \mathcal{P}_A$ be the operator that projects vectors in $\mathbb{R}^N$ into the subspace orthogonal to $\text{span}(Z_{t,-j}[A])$.

1.8.1.2 Consistency of Double Selection Lasso

We provided two critical Double Selection Lasso (DSL) assumptions in Assumption 1, to which we add the following standard DSL assumptions, adapted to the DSLFM setting. Let there exist absolute sequences $\delta_{N,T} \searrow 0$ and $\Delta_{N,T} \searrow 0$.

Assumption 5 (ASR: Approximate Sparse Regressors).

1. *Sparsity of Confounding: The confounding mapping $\delta_{t,-j}$ admits, $\forall t, j$ an exactly sparse form $||\delta_{t,-j}||_0 \leq s$.*

2. *Sparsity rate: The sparsity index obeys* $s^2 \log^2(p \vee N) / \left(\sqrt{N \log(Tp)}\right) \leq \delta_{N,T}$ *and the size of the amelioration set obeys* $\hat{s}_3 \leq C(1 \vee \hat{s}_1 \vee \hat{s}_2)$. *Additionally,* $\log^3 p/N \leq \delta_{N,T}$.

Assumption ASR(i) extends Assumption 1(ii) to include sparsity of the DSL first stage. Assumption ASR(ii) controls the rate between sparsity and the asymptotic terms p, N, T; additionally, it constrains the amelioration set to not be substantially larger than the variables selected by the LASSO procedures.

Next, we constrain the minimum and maximum m-spare eigenvalues as whenever $p > N$ the empirical design matrix $\mathbb{E}_N[z_{i,t} z_{i,t}^\top]$ will not have full rank. Define the minimal and maximal m-sparse eigenvalue of a semi-definite matrix M as

$$\phi_{\min}(m)[M] \coloneqq \min_{1 \leq ||\delta||_0 \leq m} \frac{\delta^\top M \delta}{||\delta||^2} \text{ and } \phi_{\max}(m)[M] \coloneqq \max_{1 \leq ||\delta||_0 \leq m} \frac{\delta^\top M \delta}{||\delta||^2}.$$

Assumption 6 (SE: Sparse Eigenvalues). *There exists an absolute sequence* $l_N \to \infty$ *and such that with probability of at least* $1 - \Delta_{N,T}$ *the maximal and minimal* $l_N s$*-sparse eigenvalues are bounded from above and away from zero. That is, for absolute constants* $0 < \kappa' < \kappa'' < \infty$,

$$\kappa' \leq \phi_{\min}(l_N s)[\mathbb{E}_N[z_{i,t} z_{i,t}^\top]] \leq \phi_{\max}(l_N s)[\mathbb{E}_N[z_{i,t} z_{i,t}^\top]] \leq \kappa''$$

Similarly, for $\bar{z}_i \coloneqq T^{-1} \sum_i z_{i,t}$, *we have*

$$\kappa' \leq \phi_{\min}(l_N s)[\mathbb{E}_N[\bar{z}_i \bar{z}_i^\top]] \leq \phi_{\max}(l_N s)[\mathbb{E}_N[\bar{z}_i \bar{z}_i^\top]] \leq \kappa''.$$

Requiring the minimum $m-$sparse eigenvalue to be bounded away from zero is equivalent to assuming all empirical design submatrices formed by any m components of $z_{i,t}$ are positive definite.

Next, we impose moment conditions on the structural errors and regressors.

Assumption 7 (SM: Structural Moments). *There are absolute constants* $0 < \omega < \Omega < \infty$ *and* $4 \leq \rho < \infty$ *such that for* $(y_i, \epsilon_i) \coloneqq (r_{i,t+1}, \epsilon_{i,t+1})$ *or* $(y_i, \epsilon_i) \coloneqq (z_{i,t,-j}, \epsilon^z_{i,t,j})$ *we have* $\forall i, t, j$*:*

1. $\mathbb{E}[|z_{i,t,j}|^{\rho}] \leq \Omega, \omega \leq \mathbb{E}[\epsilon_{i,t+1}^2 | z_{i,t,-j}, \epsilon_{i,t,j}^z] \leq \Omega$, *and* $\omega \leq \mathbb{E}[(\epsilon_{i,t,j}^z)^2 | z_{i,t,-j}] \leq \Omega$;

2. $\mathbb{E}[|\epsilon_i|^{\rho}] + \mathbb{E}[y_i^2] + \max_{1 \leq k \leq p}\{\mathbb{E}[z_{i,t,-j,k}^2 y_i^2] + \mathbb{E}[|z_{i,t,-j,k}^3 \epsilon_i^3|] + 1/\mathbb{E}[z_{i,t-1,-j,k}^2]\} \leq \Omega$,

3. $\max_{1 \leq k \leq p}\{\mathbb{E}[z_{i,t,-j,k}^2 \epsilon_i^2] + \mathbb{E}[z_{i,t,-j,k}^2 y_i^2]\} + \max_{1 \leq i \leq N} ||z_{i,t,-j}||_{\infty}^2 \frac{s \log(N \vee p)}{N} \leq \delta_{N,T}$ *w.p.* $1 - \Delta_{N,T}$.

4. *Weak dependence between the first- and second-stage errors: There exists a positive constant* M *such that* $\forall p, T, N$:

$$\left| \sqrt{\frac{1}{N}} \sum_{i=1}^{N} \epsilon_{i,t,j}^z \epsilon_{i,t+1} \right| \leq M \log(Tp).$$

5. *Uniformly over* t, j, *we have* $\frac{1}{N} \sum_i (\epsilon_{t,j,i}^z)^2 \xrightarrow{p} \mathbb{Z}_{t,j}^0$ *for non-stochastic real-valued scalar finite constant* $\mathbb{Z}_{t,j}^0$, *which is bounded away from zero.*

Assumptions (SM)(i)-(iii) are standard for DSL to bound various moments of the errors, characteristics, and returns. Assumption SM(iv) is novel and bounds the dependence between the first- and second-stage errors in the DSL model, which is the non-negligible asymptotic term in the DSL estimation error. This holds trivially for i.i.d. sampling in the cross-section, which we have relaxed to this specified sum. Assumption (v) is novel and introduces a uniform consistency for the second moment of the first-stage errors.

LEMMA 1. *Under the model* (1.2)*; Assumption 1; and, DSL Assumptions ASR, SE, and SM, the DSL estimator has the property that*

$$\max_{t,j} |\widehat{c}_{t+1,j} - c_{t+1,j}| = O_p\left(\sqrt{\frac{\log(Tp)}{N}}\right).$$

Proof of Lemma A1. We proceed with the decomposition of the estimation error using the definition of the DSL estimator and model (1.2).

$$
\begin{aligned}
\hat{c}_{t+1,j} - c_{t+1,j} &= \left(z_{t,j}^\top \mathcal{M}_{\hat{I}} z_{t,j}\right)^{-1} \left(z_{t,j}^\top \mathcal{M}_{\hat{I}} (Z_{t,-j} c_{t+1,-j} + \epsilon_{t+1})\right) \\
&= \left(z_{t,j}^\top \mathcal{M}_{\hat{I}} z_{t,j}\right)^{-1} \left(Z_{t,-j}\delta_{t,j}\right)^\top \mathcal{M}_{\hat{I}} Z_{t,-j} c_{t+1,-j} \\
&\quad + \left(z_{t,j}^\top \mathcal{M}_{\hat{I}} z_{t,j}\right)^{-1} \left(Z_{t,-j}\delta_{t,j}\right)^\top \mathcal{M}_{\hat{I}} \epsilon_{t+1} \\
&\quad + \left(z_{t,j}^\top \mathcal{M}_{\hat{I}} z_{t,j}\right)^{-1} \epsilon_{t,j}^{z\top} \mathcal{M}_{\hat{I}} Z_{t,-j} c_{t+1,-j} \\
&\quad - \left(z_{t,j}^\top \mathcal{M}_{\hat{I}} z_{t,j}\right)^{-1} \epsilon_{t,j}^{z\top} \mathcal{P}_{\hat{I}} \epsilon_{t+1} \\
&\quad + \left(z_{t,j}^\top \mathcal{M}_{\hat{I}} z_{t,j}\right)^{-1} \epsilon_{t,j}^{z\top} \epsilon_{t+1}.
\end{aligned}
$$

From Belloni, Chernozhukov, and Hansen (2014) under the aforementioned DSL assumptions, the last term in this five-term decomposition is the asymptotically relevant term while the remaining terms are asymptotically negligible. We first handle the denominator of the fifth term before dealing with the entire term.

$$
\begin{aligned}
N^{-1} z_{t,j}^\top \mathcal{M}_{\hat{I}} z_{t,j} &= N^{-1} \left(Z_{t,-j}\delta_{t,j} + \epsilon_{t,j}^z\right)^\top \mathcal{M}_{\hat{I}} \left(Z_{t,-j}\delta_{t,j} + \epsilon_{t,j}^z\right) \\
&= \epsilon_{t,j}^{z\top}\epsilon_{t,j}^z/N + \delta_{t,j}^\top Z_{t,-j}^\top \mathcal{M}_{\hat{I}} Z_{t,-j}\delta_{t,j}/N + 2\delta_{t,j}^\top Z_{t,-j}^\top \mathcal{M}_{\hat{I}} \epsilon_{t,j}^z/N - \epsilon_{t,j}^{z\top} \mathcal{P}_{\hat{I}} \epsilon_{t,j}^z/N \\
&\lesssim_P \epsilon_{t,j}^{z\top}\epsilon_{t,j}^z/N + o_p(1)
\end{aligned}
$$

where the first equality holds by definition of the first-stage; the second equality holds by multiplying out the terms and by definition of the projection matrices; and, the probabilistic bound holds given the latter three terms are asymptotically negligible, as in the proof of Theorem 1 in Belloni, Chernozhukov, and Hansen (2014), as compared to the sum of second moments of the first-stage errors. Thus, by Assumption SM(v), we conclude $\epsilon_{t,j}^{z\top}\epsilon_{t,j}^z/N$ converges in probability uniformly over t, j, to $\mathbb{Z}_{t,j}^0$.

We proceed with the uniform consistency result.

$$\begin{aligned}\max_{t,j}|\widehat{c}_{t+1,j}-c_{t+1,j}| &\lesssim_P \max_{t,j}\left|\left(N^{-1}z_{t,j}^\top \mathcal{M}_{\hat{I}} z_{t,j}\right)^{-1}\frac{1}{N}\sum_{i=1}^{N}\epsilon_{i,t,j}^z\epsilon_{i,t+1}\right| \\ &\lesssim_P \sqrt{\frac{1}{N}}\max_{t,j}\left|\sqrt{\frac{1}{N}}\sum_{i=1}^{N}\epsilon_{i,t,j}^z\epsilon_{i,t+1}\right| \\ &\lesssim_P \sqrt{\frac{\log(Tp)}{N}}\end{aligned}$$

which holds for the first probabilistic bound by substituting the decomposition above; for the second probabilistic bound, using the above result to replace the denominator with a constant that is bounded away from zero uniformly over t, j; and, the final bound holds by assumption SM(iv). In the case of i.i.d. sampling in the cross-section or if the dependence is sufficiently weak such that SM(iv) holds, then we can invoke Lemma A.4 in Belloni et al. (2018) to conclude the mean zero scalar random variable $\epsilon_{i,t,j}^z\epsilon_{i,t+1}$, given the moment conditions of the DSL model, has a maximal deviation that converges in probability to zero at the specified rate if we further constrain the moment of the mean zero random variable $\mathbb{E}\left[\max_{t,j}\left|\epsilon_{i,t,j}^z\epsilon_{i,t+1}\right|^q\right] \leq M^q$ for $q > 2$ and absolute constant M uniformly across t, j. □

1.8.1.3 Consistency of Latent Factors and Loadings

We first prove a bound on the distance between the infeasible and feasible symmetric matrix used in the eigendecompositions. Let $\widehat{\Lambda}_{Tp} \in \mathbb{R}^{k\times k}$ be a diagonal matrix containing the k largest eigenvalues of $(Tp)^{-1}\widehat{C}\widehat{C}^\top$ and similarly for $\Lambda_{Tp} \in \mathbb{R}^{k\times k}$, a diagonal matrix containing the k largest eigenvalues of $(Tp)^{-1}\widehat{C}\widehat{C}^\top$.

LEMMA 2. *Under the assumptions of Lemma A1,* $\left\|(Tp)^{-1}\widehat{C}\widehat{C}^\top - (Tp)^{-1}CC^\top\right\| = O_p\left(\frac{\log Tp}{N}\right)$.

Proof of Lemma A2.

$$\begin{aligned}
\left\|\widehat{C}\widehat{C}^\top - CC^\top\right\| &= \left\|\widehat{C}\widehat{C}^\top - C\widehat{C}^\top + C\widehat{C}^\top - CC^\top\right\| \\
&\leq \left\|C\widehat{C}^\top - CC^\top\right\| + \left\|\widehat{C}\widehat{C}^\top - C\widehat{C}^\top\right\| \\
&\leq \|C\| \left\|\widehat{C} - C\right\| + \left\|\widehat{C} - C\right\| \left\|\widehat{C}\right\| \\
&= \|C\| \left\|\widehat{C} - C\right\| + \left\|\widehat{C} - C\right\| \left\|\widehat{C} - C + C\right\| \\
&\leq \|C\| \left\|\widehat{C} - C\right\| + \left\|\widehat{C} - C\right\| \left(\left\|\widehat{C} - C\right\| + \|C\|\right) \\
&= 2\|C\| \left\|\widehat{C} - C\right\| + \left\|\widehat{C} - C\right\|^2 \\
&\lesssim_P \sqrt{sT} \left\|\widehat{C} - C\right\| + \left\|\widehat{C} - C\right\|^2 \\
&\leq \sqrt{spT^2} \max_{t,j} |\widehat{c}_{t+1,j} - c_{t+1,j}| + Tp \max_{t,j} |\widehat{c}_{t+1,j} - c_{t+1,j}|^2 \\
&\lesssim_P \sqrt{\frac{spT^2 \log(Tp)}{N}} + \frac{Tp \log(Tp)}{N} \lesssim \frac{Tp \log(Tp)}{N}.
\end{aligned}$$

where the first and third inequality holds by the triangle inequality; the second inequality holds by Cauchy-Schwarz; the first probabilistic bound holds by Assumption 1(i) / 5(i); and, the last probabilistic bound holds by Lemma A1. We use the final bound for simplicity. □

We next bound the estimation error between the feasible and infeasible eigenvalues.

LEMMA 3. *Under the assumptions of Lemma A2 and Assumption 3,*

$$\left\|\widehat{\Lambda}_{Tp} - \Lambda_{Tp}\right\|^2 = O_p\left(\frac{s^2 \log^2(Tp)}{N^2}\right).$$

Proof of Lemma A3.

$$\begin{aligned}
\left\|\widehat{\Lambda}_{Tp} - \Lambda_{Tp}\right\|^2 &= \sum_{l=1}^{k} \sum_{i=1}^{k} (\widehat{\lambda}_l - \lambda_i)^2 \\
&\leq k^2 \max_{l \in \{1,\ldots,k\}} |\widehat{\lambda}_l - \lambda_l|^2 \\
&\leq k^2 \max_{l \in \{1,\ldots,T\}} |\widehat{\lambda}_l - \lambda_l|^2 \\
&\leq \frac{k^2}{T^2 p^2} \left\|\widehat{C}\widehat{C}^\top - CC^\top\right\|^2 = O_p(\frac{s^2 \log^2(Tp)}{N^2})
\end{aligned}$$

where the first equality is the definition of the Frobenius norm; the first inequality bounds the sum by the maximum element; the second inequality bounds the maximum deviation between the k largest eigenvalues of the feasible and infeasible decompositions by the deviations between all T eigenvalues; the last inequality controls the stability of the spectrum by applying Weyl's inequality from Theorem 4.5.3 of Vershynin (2018); and, the probabilistic bound holds by Lemma A2. □

We next prove a lemma for the time series average of the ℓ_2 norm of the feasible and infeasible eigenvectors.

LEMMA 4. *Under Assumption 2(iii) and those of Lemma A3, there exists an orthogonal matrix* $\widehat{O} \in \mathbb{R}^{k \times k}$ *such that* $\left\| \widehat{F} - \widetilde{F}\widehat{O}^\top \right\|^2 = O_p\left(\frac{\log^2(Tp)}{N^2}\right)$.

Proof of Lemma A4. We use a variant of the Davis-Kahan theorem shown in Yu, Wang, and Samworth (2015) where

$$\delta \coloneqq \min_{i:i \neq l} |\lambda_l - \lambda_i| > 0,$$

which holds by Assumption 2(iii), to conclude for some $\widehat{O} \in \mathbb{R}^{k \times k}$ orthogonal matrix that

$$2^{3/2}\delta^{-1}(Tp)^{-2}\left\|\widehat{C}\widehat{C}^\top - CC^\top\right\|^2 \geq ||\widehat{F}\widehat{O} - \widetilde{F}||^2 = ||\widehat{F} - \widetilde{F}\widehat{O}^\top||^2$$

where the inequality is the use of the variant of the Davis-Kahan theorem, bounding the distance between the eigenvectors by the distance between the original matrices, and the equality follows given post multiplying by an orthogonal matrix does not change the Frobenius norm. The rate in the result then follows given Lemma A2. □

We next bound the ℓ_2 norm between the feasible and infeasible eigenvectors.

LEMMA 5. *Under the assumptions of Lemma A4,* $\left\| \widehat{f}_{t+1} - \widehat{O}^\top \widetilde{f}_{t+1} \right\| = O_p\left(\sqrt{\frac{s\log(Tp)}{N}}\right)$.

Proof of Lemma A5. First, we perform the following decomposition using the definition of the eigenvectors.

$$\begin{aligned}
\widehat{f}_{t+1} - \widehat{O}^\top \widetilde{f}_{t+1} &= (Tp)^{-1}\widehat{\Lambda}_{Tp}^{-1}\widehat{F}^\top \widehat{C}\widehat{C}_{t+1} - (Tp)^{-1}\widehat{O}^\top \Lambda_{Tp}^{-1}\widetilde{F}^\top CC_{t+1} \\
&= (Tp)^{-1}\left(\widehat{\Lambda}_{Tp}^{-1} - \Lambda_{Tp}^{-1}\right)\widehat{F}^\top\left(\widehat{C}\widehat{C}_{t+1} - CC_{t+1}\right) \\
&\quad + (Tp)^{-1}\widehat{O}^\top\Lambda_{Tp}^{-1}\left(\widetilde{F}^\top - \widehat{F}^\top\right)CC_{t+1} \\
&\quad + (Tp)^{-1}\left(\widehat{\Lambda}_{Tp}^{-1} - \widehat{O}^\top\Lambda_{Tp}^{-1}\right)\widehat{F}^\top CC_{t+1} \\
&\quad + (Tp)^{-1}\Lambda_{Tp}^{-1}\widehat{F}^\top\left(\widehat{C}\widehat{C}_{t+1} - CC_{t+1}\right)
\end{aligned}$$

where the equality follows by adding and subtracting terms.

Thus,

$$\begin{aligned}
\left\|\widehat{f}_{t+1} - \widehat{O}^\top \widetilde{f}_{t+1}\right\| &\leq (Tp)^{-1}\left\|\widehat{\Lambda}_{Tp}^{-1} - \Lambda_{Tp}^{-1}\right\|\left\|\widehat{F}\right\|\left\|\widehat{C}\widehat{C}_{t+1} - CC_{t+1}\right\| \\
&\quad + (Tp)^{-1}\left\|\Lambda_{Tp}^{-1}\right\|\left\|\widehat{F} - \widetilde{F}\right\|\|CC_{t+1}\| \\
&\quad + (Tp)^{-1}\left\|\widehat{\Lambda}_{Tp}^{-1} - \Lambda_{Tp}^{-1}\right\|\left\|\widehat{F}\right\|\|CC_{t+1}\| \\
&\quad + (Tp)^{-1}\left\|\Lambda_{Tp}^{-1}\right\|\left\|\widehat{F}\right\|\left\|\widehat{C}\widehat{C}_{t+1} - CC_{t+1}\right\| \\
&= O_p\left(\sqrt{\frac{s^3\log(Tp)^3}{pN^3}}\right) + O_p\left(\frac{\sqrt{T}s\log(Tp)}{TpN}\right) \\
&\quad + O_p\left(\frac{s^2\log(Tp)}{pN}\right) + O_p\left(\sqrt{\frac{s\log(Tp)}{N}}\right) \\
&= O_p\left(\sqrt{\frac{s\log(Tp)}{N}}\right)
\end{aligned}$$

where the first inequality follows from the aforementioned decomposition in this proof with the use of the triangle and Cauchy-Schwarz inequalities; the first probabilistic bound holds given the normalization that $\widehat{F}^\top\widehat{F}/T = I_k$ then $||\widehat{F}|| = \sqrt{Tk}$, given Λ_{Tp} contains k nonzero real-valued eigenvalues bounded away from zero by Assumption 2(iii), given the rates from Lemmas A3 and A4 (which gives same rate by CMT), given—similar to lemma A2—$||\widehat{C}\widehat{C}_{t+1} - CC_{t+1}|| = O_p(\sqrt{\frac{sTp^2\log(Tp)}{N}})$, and $||CC_{t+1}|| = O_p(\sqrt{s^2T})$ by Assumption 1(ii); and, the final probabilistic bound holds for simplicity of exposition. □

LEMMA 6. *Under the assumptions of A5, for* $H^\top = \widehat{O}^\top \Lambda_{Tp}^{-1}(F^\top F^0/T)(\Gamma_\beta^{0\top}\Gamma_\beta^0/p)$ *we have*

$$\widehat{O}\widetilde{f}_{t+1} - H^\top f_{t+1}^0 = 0.$$

Proof of Lemma A6.

$$\begin{aligned}
\widehat{O}\widetilde{f}_{t+1} &= (Tp)^{-1}\widehat{O}^\top \Lambda_{Tp}^{-1} F^\top C C_{t+1} \\
&= (Tp)^{-1}\widehat{O}^\top \Lambda_{Tp}^{-1} F^\top F \Gamma_\beta^\top \Gamma_\beta f_{t+1} \\
&= (Tp)^{-1}\widehat{O}^\top \Lambda_{Tp}^{-1} F^\top F^0 \Gamma_\beta^{0\top}\Gamma_\beta^{0\top} f_{t+1}^0 \\
&= H^\top f_{t+1}^0.
\end{aligned}$$

where the first equality holds by the definition of the infeasible eigenvectors; the second equality holds given the definition of C; the third equality holds given the definitions of the true loadings and factors as rotations of the observed ones; and, the final equality holds by definition of the H matrix. □

Finally, using the above lemmas, we prove Propositions 1 and 2.

Proof of Proposition 1.

$$\begin{aligned}
\left\|\widehat{f}_{t+1} - H^\top f_{t+1}^0\right\| &\leq \left\|\widehat{f}_{t+1} - \widehat{O}^\top \widetilde{f}_{t+1}\right\| + \left\|\widehat{O}^\top \widetilde{f}_{t+1} - H^\top f_{t+1}^0\right\| \\
&= O_p\left(\sqrt{\frac{s\log(Tp)}{N}}\right)
\end{aligned}$$

by Lemmas A5 and A6. □

Next, we provide a lemma for the ℓ_∞ norm of the PCA estimation error for the loadings, which will allow us to obtain norms on the soft-threshold estimation error for the loadings by use of a tool from the high-dimensional econometrics handbook (Belloni et al. 2018).

LEMMA 7. *Under the assumptions of Lemma A1 and Assumption 3,*

$$\left\|\widehat{\Gamma}_\beta - \Gamma_\beta^0 (H^\top)^{-1}\right\|_\infty = O_p\left(\sqrt{\frac{\log(Tp)}{N}}\right).$$

Proof of Lemma A7.

$$\begin{aligned}
\left\|\widehat{\Gamma}_\beta - \Gamma^0_\beta (H^\top)^{-1}\right\|_\infty &= \max_j \left(\sum_{l=1}^k \left| \widehat{\Gamma}_{\beta,j,l} - (\Gamma^0_{\beta,j})^\top (H^\top)^{-1}_l \right| \right) \\
&= \max_j \left\| \widehat{\Gamma}_{\beta,j} - H^{-1}\Gamma^0_{\beta,j} \right\|_1 \\
&= \max_j \left\| \widehat{\Gamma}_{\beta,j} \pm \widetilde{\Gamma}_{\beta,j} - H^{-1}\Gamma^0_{\beta,j} \right\|_1 \\
&\leq \max_j \left\| \widehat{\Gamma}_{\beta,j} - \widetilde{\Gamma}_{\beta,j} \right\|_1 + \max_j \left\| \widetilde{\Gamma}_{\beta,j} - H^{-1}\Gamma^0_{\beta,j} \right\|_1 \\
&= \max_j \left\| T^{-1}\widehat{F}^\top \widehat{C}_j - T^{-1}F^\top C_j \right\|_1 + \max_j \left\| T^{-1}F^\top C_j - H^{-1}\Gamma^0_{\beta,j} \right\|_1 \\
&= T^{-1} \max_j \left\| \widehat{F}^\top \widehat{C}_j \pm \widehat{F}^\top C_j - F^\top C_j \right\|_1 + \max_j \left\| T^{-1}F^\top F\Gamma_{\beta,j} - H^{-1}\Gamma^0_{\beta,j} \right\|_1 \\
&\leq T^{-1} \max_j \left\| \widehat{F}^\top \left(\widehat{C}_j - C_j \right) \right\|_1 + T^{-1} \max_j \left\| \left(\widehat{F} - F \right)^\top C_j \right\|_1 \\
&\quad + \max_j \left\| T^{-1}F^\top F\Gamma_{\beta,j} - H^{-1}\Gamma^0_{\beta,j} \right\|_1 \\
&\leq \frac{\sqrt{k}}{T} \left\| \widehat{F} \right\|_2 \max_j \left\| \widehat{C}_j - C_j \right\|_2 + \frac{\sqrt{k}}{T} \left\| \widehat{F} - F \right\|_2 \max_j \left\| C_j \right\|_2 \\
&\quad + \sqrt{k} \left\| T^{-1}F^\top F - I_{k\times k} \right\|_2 \left\| H^{-1} \right\|_2 \max_j \left\| \Gamma^0_{\beta,j} \right\|_2 \\
&\lesssim_P \max_{t,j} |\widehat{c}_{t+1,j} - c_{t+1,j}| + T^{-1/2} \left\| \widehat{F} - F \right\| \\
&\lesssim_P O_p\left(\sqrt{\frac{\log(Tp)}{N}} \right) + O_p\left(\frac{\log(Tp)}{N} \right)
\end{aligned}$$

where the first two equalities follow from the definition of the ℓ_∞ norm; the third equality adds and subtracts; the first inequality uses the triangle inequality; the fourth equality uses the definitions of the feasible and infeasible estimator; the second inequality uses a triangle inequality; the last inequality uses $\left\|A^\top x\right\|_1 \leq \sqrt{k}\left\|A\right\|_2 \left\|x\right\|_2$; the first probabilistic bound uses $\left\|\widehat{F}\right\|_2 = O_p(\sqrt{kT})$, the sum of the estimation errors in the characteristic portfolios, $|\widehat{c}_{t+1,j} - c_{t+1,j}|$ is bounded by T times the maximum element, the elements of C_j are bounded random variables hence $\max_j \|C_j\| = O_p(\sqrt{T})$, $T^{-1}\sum_t f^0_{t+1} f^{0,\top}_{t+1} \to_p \Sigma_f$ by Assumption 2(i), $\|H^{-1}\| = O_p(1)$ by definition of the invertible H matrix, and Γ^0_β contains bounded elements uniformly over j by Assumption 2(ii); and, the last probabilistic bound follows by Lemmas

A1 and A4 and assumption 5(ii). □

Proof of Proposition 2. In Lemma A7, we show $\left\|\widehat{\Gamma}_\beta - \Gamma_\beta^0 (H^\top)^{-1}\right\|_\infty \lesssim_P \sqrt{\frac{\log(Tp)}{N}}$, which allows us to invoke Theorem 2.10 from Belloni et al. (2018) under exact sparsity of Γ_β^0 where λ, the hyperparameter used to soft-threshold the ℓ_1 norm of the rows of $\widehat{\Gamma}_\beta$, is selected such that with probability approaching 1,

$$\lambda \geq (1-\alpha) - \text{quantile of } \left\|\widehat{\Gamma}_\beta - \Gamma_\beta^0 (H^\top)^{-1}\right\|_\infty.$$

That is, λ can be set to the product of a large constant and the rate of the ℓ_∞ norm, $\sqrt{\frac{\log(Tp)}{N}}$, to ensure this inequality holds.[12] Then, by Theorem 2.10 given $\alpha \to 0$ and $\lambda \lesssim \sqrt{\log(Tp)/N}$, we have for all $q \geq 1$

$$\left\|\check{\Gamma}_{\beta,l} - \Gamma_\beta^0 (H^\top)_l^{-1}\right\|_q \lesssim_P s^{1/q}\sqrt{\frac{\log(Tp)}{N}}.$$

This holds column-by-column for the matrix estimation error $\check{\Gamma}_\beta - \Gamma_\beta^0$, which we can thus square and sum together for the squared Frobenius norm of the estimation error at the same rate. □

1.8.1.4 Consistency and Normality of Observable Factor Risk Premia

In these results, we are using the eigenvectors and loadings derived from the demeaned characteristic portfolio matrix: $\widehat{C}^D \coloneqq \widehat{C} - \iota_T T^{-1}\sum_t \widehat{c}_{t+1}^\top$. That is, $\widehat{V}$ are the $\sqrt{T}$ scaled eigenvectors associated with the k largest eigenvalues of $(Tp)^{-1}\widehat{C}^D\widehat{C}^{D\top}$. Further, $\widehat{\Gamma}_\beta^D = T^{-1}\widehat{C}^{D\top}\widehat{V}$. The results established in the above subsection would follow analogously for this new notation as we simply have mean zero eigenvectors. Finally, our cross sectional and time series OLS estimators are standard:

$$\widehat{\gamma} \coloneqq \left(\bar{\widehat{\beta}}^\top \bar{\widehat{\beta}}\right)^{-1} \bar{\widehat{\beta}}^\top \bar{r}, \qquad \widehat{\eta} \coloneqq \left(\widehat{V}^\top \widehat{V}\right)^{-1} \widehat{V}^\top G,$$

where $\bar{\widehat{\beta}} \coloneqq \bar{Z}\check{\Gamma}_\beta^D$ for $\bar{Z} = T^{-1}\sum_t Z_t$ for $Z_t \in \mathbb{R}^{N\times p}$, $\forall t$. The same holds for the time series average return $\bar{r} \in \mathbb{R}^N$.

[12] In practice, we cross-validate for a finite-sample optimal λ.

LEMMA 8. *Under the models (1.2) and (1.4); Assumptions 1, 2, and 3; and, DSL Assumptions ASR, SE, and SM; we have*

$$\left\| N^{-1}\bar{\widehat{\beta}}^{\top}\bar{\widehat{\beta}} - N^{-1}\bar{\beta}^{\top}\bar{\beta} \right\| = O_p\left(\sqrt{\frac{s^2\log(Tp)}{N}}\right).$$

Proof of Lemma A8.

$$\begin{aligned}
\left\| N^{-1}\bar{\widehat{\beta}}^{\top}\bar{\widehat{\beta}} - N^{-1}\bar{\beta}^{\top}\bar{\beta} \right\| &= N^{-1}\left\| \bar{\widehat{\beta}}^{\top}\bar{\widehat{\beta}} \pm \bar{\widehat{\beta}}^{\top}\bar{\beta} - \bar{\beta}^{\top}\bar{\beta} \right\| \\
&\leq N^{-1}\left\| \bar{\widehat{\beta}}^{\top}\bar{\widehat{\beta}} - \bar{\widehat{\beta}}^{\top}\bar{\beta} \right\| + N^{-1}\left\| \bar{\widehat{\beta}}^{\top}\bar{\beta} - \bar{\beta}^{\top}\bar{\beta} \right\| \\
&= N^{-1}\left\| \left(\bar{\widehat{\beta}} \pm \bar{\beta}\right)^{\top}\left(\bar{\widehat{\beta}} - \bar{\beta}\right) \right\| + N^{-1}\left\| \left(\bar{\widehat{\beta}} - \bar{\beta}\right)^{\top}\bar{\beta} \right\| \\
&\leq N^{-1}\left\| \bar{\widehat{\beta}} - \bar{\beta} \right\|^2 + \frac{2}{N}\left\| \left(\bar{\widehat{\beta}} - \bar{\beta}\right)^{\top}\bar{\beta} \right\| \\
&= N^{-1}\left\| \bar{Z}\left(\check{\Gamma}^D_\beta - \Gamma^0_\beta\right) \right\|^2 + \frac{2}{N}\left\| \left(\check{\Gamma}^D_\beta - \Gamma^0_\beta\right)^{\top}\bar{Z}^{\top}\bar{Z}\Gamma^0_\beta \right\| \\
&\leq \left\| \left(\check{\Gamma}^D_\beta - \Gamma^0_\beta\right)^{\top}\frac{\bar{Z}^{\top}\bar{Z}}{N}\left(\check{\Gamma}^D_\beta - \Gamma^0_\beta\right) \right\|^2 \\
&\quad + 2\left\| \left(\check{\Gamma}^D_\beta - \Gamma^0_\beta\right)^{\top}\frac{\bar{Z}^{\top}\bar{Z}}{N}\Gamma^0_\beta \right\| \\
&\leq \left\| \check{\Gamma}^D_{\beta,l} - \Gamma^0_{\beta,l} \right\|^2 \left\| \check{\Gamma}^D_\beta - \Gamma^0_\beta \right\|^2 \phi^2_{\max}(2s)\left[\frac{\bar{Z}^{\top}\bar{Z}}{N}\right] \\
&\quad + 2\left\| \check{\Gamma}^D_{\beta,l} - \Gamma^0_{\beta,l} \right\| \left\| \Gamma^0_\beta \right\| \phi_{\max}(2s)\left[\frac{\bar{Z}^{\top}\bar{Z}}{N}\right] \\
&\lesssim_P O_p\left(\frac{s^2\log^2(Tp)}{N^2}\right) + O_p\left(\sqrt{\frac{s^2\log(Tp)}{N}}\right)
\end{aligned}$$

where the first equality follows by adding and subtracting; the first inequality follows by the triangle inequality; the second equality follows from adding and subtracting and rearranging; the second inequality follows from the triangle and Cauchy-Schwartz inequalities; the third equality follows from the definition of the factor loading estimator and the average factor loading; the third inequality follows from $||A|| \leq ||AA\top||$; the final inequality follows from multiplying and dividing by the norms to obtain unit vectors and then bounding with the $s+\hat{s}$-maximally sparse eigenvalue of $\frac{\bar{Z}\top\bar{Z}}{N}$ (where the $s+\hat{s} \leq 2s$ by thresholding estimator); and,

the probabilistic bound holds given the maximum eigenvalue is bounded by DSL Assumption SE, $\left\|\Gamma_\beta^0\right\| = O_p(\sqrt{s})$ by Assumption 1(ii), and $\left\|\check{\Gamma}_{\beta,l}^D - \Gamma_{\beta,l}^0\right\| = O_p\left(\sqrt{\frac{s\log(Tp)}{N}}\right)$, which is the same rate for the entire matrix per Proposition 2. □

LEMMA 9. *Under the assumptions of Lemma A8, we have*

$$\frac{\sqrt{T}}{N}\left\|\hat{\bar{\beta}}^\top \bar{r} - \bar{\beta}^\top \bar{r}\right\| = O_p\left(\sqrt{\frac{Ts^2\log(Tp)}{N}}\right).$$

Proof of Lemma A9.

$$\begin{aligned}
\frac{\sqrt{T}}{N}\left\|\hat{\bar{\beta}}^\top \bar{r} - \bar{\beta}^\top \bar{r}\right\| &= \frac{\sqrt{T}}{N}\left\|\left(\check{\Gamma}_\beta^D - \Gamma_\beta^0\right)^\top \bar{Z}^\top\left(\bar{Z}\Gamma_\beta^0\gamma^0 + T^{-1}\sum_t Z_t\Gamma_\beta^0 v_{t+1}^0 + \bar{\epsilon}\right)\right\| \\
&\leq \frac{\sqrt{T}}{N}\left\|\left(\check{\Gamma}_\beta^D - \Gamma_\beta^0\right)^\top \bar{Z}^\top\bar{Z}\Gamma_\beta^0\gamma^0\right\| \\
&\quad + \frac{\sqrt{T}}{N}\left\|\left(\check{\Gamma}_\beta^D - \Gamma_\beta^0\right)^\top \bar{Z}^\top T^{-1}\sum_t Z_t\Gamma_\beta^0 v_{t+1}^0\right\| \\
&\quad + \frac{\sqrt{T}}{N}\left\|\left(\check{\Gamma}_\beta^D - \Gamma_\beta^0\right)^\top \bar{Z}^\top\bar{\epsilon}\right\| \\
&\leq \sqrt{T}\left\|\check{\Gamma}_{\beta,l}^D - \Gamma_{\beta,l}^0\right\|\left\|\Gamma_\beta^0\gamma^0\right\|\phi_{\max}(2s)\left[\frac{\bar{Z}^\top\bar{Z}}{N}\right] \\
&\quad + \left\|\check{\Gamma}_{\beta,l}^D - \Gamma_{\beta,l}^0\right\|^2\phi_{\max}(2s)\left[\frac{\bar{Z}^\top\bar{Z}}{N}\right]\sqrt{\frac{T}{N}}\left\|T^{-1}\sum_t Z_t\Gamma_\beta^0 v_{t+1}^0\right\| \\
&\quad + \left\|\check{\Gamma}_{\beta,l}^D - \Gamma_{\beta,l}^0\right\|^2\phi_{\max}(2s)\left[\frac{\bar{Z}^\top\bar{Z}}{N}\right]\sqrt{\frac{T}{N}}\left\|\bar{\epsilon}\right\| \\
&\lesssim_p O_p\left(\sqrt{\frac{Ts^2\log(Tp)}{N}}\right) + O_p\left(\frac{s^{3/2}\log(Tp)}{N}\right) \\
&\quad + O_p\left(\frac{s\log(Tp)}{N}\right)
\end{aligned}$$

where the first equality follows by the definitions of the factor loading estimator, the average factor loading, and the time series average return of assets; the first inequality follows by the triangle inequality; the second inequality holds by an analogous argument to Lemma A8; and, the probabilistic bound holds again by an analogous argument to Lemma A8 with the additional bounds of $\left\|\Gamma_\beta^0\gamma^0\right\| = O_p(\sqrt{s})$ by Assumptions 2(i)-(ii) and 1(ii),

$\left\|T^{-1}\sum_t Z_t\Gamma_\beta^0 v_{t+1}^0\right\| = O_p(\sqrt{\frac{sN}{T}})$ by Assumption 4(ii), and $\|\bar{\epsilon}\| = O_p(\sqrt{\frac{N}{T}})$ by Assumption 4(i). □

LEMMA 10. *Under the assumptions of Lemma A8 and Assumption 4(iii), we have*

$$\sqrt{T}\,(\widehat{\gamma} - H\gamma_0) = O_p(1).$$

Proof of Lemma A10. We decompose the estimation error into three terms.

$$\begin{aligned}
\sqrt{T}\,(\widehat{\gamma} - H\gamma_0) &= \sqrt{T}\,(\widehat{\gamma} - \widetilde{\gamma}) + \sqrt{T}\,(\widetilde{\gamma} - H\gamma_0) \\
&= \left(\frac{\bar{\widehat{\beta}}^\top \bar{\widehat{\beta}}}{N}\right)^{-1} \frac{\sqrt{T}\bar{\widehat{\beta}}^\top \bar{r}}{N} \pm \left(\frac{\bar{\beta}^\top \bar{\beta}}{N}\right)^{-1} \frac{\sqrt{T}\bar{\beta}^\top \bar{r}}{N} - \sqrt{T}H\gamma_0 \\
&= \underbrace{\left(\frac{\bar{\widehat{\beta}}^\top \bar{\widehat{\beta}}}{N}\right)^{-1} \frac{\sqrt{T}\bar{\widehat{\beta}}^\top \bar{r}}{N} - \left(\frac{\bar{\beta}^\top \bar{\beta}}{N}\right)^{-1} \frac{\sqrt{T}\bar{\beta}^\top \bar{r}}{N}}_{\mathcal{A}_\gamma} \\
&\quad + \underbrace{\sqrt{T}H^\top \left(\frac{\Gamma_\beta^{0\top}\bar{Z}^\top \bar{Z}\Gamma_\beta^0}{N}\right)^{-1} \frac{\Gamma_\beta^{0\top}\bar{Z}^\top}{N} \frac{1}{T}\sum_t Z_t\Gamma_\beta^0 v_{t+1}^0}_{\mathcal{B}_\gamma} \\
&\quad + \underbrace{\sqrt{T}H^\top \left(\frac{\Gamma_\beta^{0\top}\bar{Z}^\top \bar{Z}\Gamma_\beta^0}{N}\right)^{-1} \frac{\Gamma_\beta^{0\top}\bar{Z}^\top}{N}\bar{\epsilon}}_{\mathcal{C}_\gamma} \\
&=: \mathcal{A}_\gamma + \mathcal{B}_\gamma + \mathcal{C}_\gamma.
\end{aligned}$$

where the first equality follows by adding and subtracting; the second equality follows from the definition of the feasible and infeasible estimator; the last equality follows from the definition of the time series average of asset returns and rearranging; and, finally, we define three terms to prove, in the rest of this proof, $\mathcal{A}_\gamma + \mathcal{C}_\gamma = o_p(1)$ and $\mathcal{B}_\gamma = O_p(1)$ to prove the lemma.

First, we prove $\mathcal{A}_\gamma = o_P(1)$. Define notation:

$$\mathcal{A}_\gamma \coloneqq \widehat{\mathcal{A}}_i^{-1}\widehat{\mathcal{A}}_{ii} - \mathcal{A}_i^{-1}\mathcal{A}_{ii} \coloneqq \left(\frac{\bar{\widehat{\beta}}^\top \bar{\widehat{\beta}}}{N}\right)^{-1} \frac{\sqrt{T}\bar{\widehat{\beta}}^\top \bar{r}}{N} - \left(\frac{\bar{\beta}^\top \bar{\beta}}{N}\right)^{-1} \frac{\sqrt{T}\bar{\beta}^\top \bar{r}}{N}$$

and $\Delta_i := \widehat{\mathcal{A}}_i^{-1} - \mathcal{A}_i^{-1}$ and $\Delta_{ii} := \widehat{\mathcal{A}}_{ii} - \mathcal{A}_{ii}$. Thus,

$$\widehat{\mathcal{A}}_i^{-1}\widehat{\mathcal{A}}_{ii} - \mathcal{A}_i^{-1}\mathcal{A}_{ii} = \mathcal{A}_{ii}\Delta_i + \mathcal{A}_i^{-1}\Delta_{ii} + \Delta_i\Delta_{ii} = O_p(1)o_p(1) + O_p(1)o_p(1) + o_p(1)o_p(1) = o_p(1)$$

given: by Lemma A9, $\mathcal{A}_{ii} = O_p(1)$ and $\Delta_{ii} = o_p(1)$; and, by Lemma A8 and CMT, $\mathcal{A}_i = O_p(1)$ and $\Delta_i = o_p(1)$.

Second, $\mathcal{B}_\gamma = O_p(1)$ given

$$\begin{aligned}
\|\mathcal{B}_\gamma\| &= \sqrt{T}\left\| H^\top \left(\frac{\Gamma_\beta^{0\top}\bar{Z}^\top\bar{Z}\Gamma_\beta^0}{N} \right)^{-1} \frac{\Gamma_\beta^{0\top}\bar{Z}^\top}{N} \frac{1}{T}\sum_t Z_t\Gamma_\beta^0 v_{t+1}^0 \right\| \\
&\leq \|H\| \left\| \left(\frac{\Gamma_\beta^{0\top}\bar{Z}^\top\bar{Z}\Gamma_\beta^0}{N} \right)^{-1} \right\|_2 \left\| \frac{\Gamma_\beta^{0\top}\bar{Z}^\top}{\sqrt{N}} \right\| \left\| \sqrt{\frac{1}{NT}}\sum_t Z_t\Gamma_\beta^0 v_{t+1}^0 \right\| \\
&\lesssim_P \phi_{\min}^{-1}(2s)[N^{-1}\bar{Z}^\top\bar{Z}]
\end{aligned}$$

where the first equality substitutes the notation; the first inequality follows by $\|ABx\| \leq \|A\|\,\|B\|_2\,\|x\|$ for matrices A, B and vector x where $\|B\|_2$ is the spectral norm; and, the probabilistic bound follows given $\|H\| = O_p(1)$, the spectral norm of the inverse matrix is bounded by the $2s$-sparse minimum eigenvalue of $N^{-1}\bar{Z}^\top\bar{Z}$ which is bounded away from zero by DSL Assumption SE, $\left\|\bar{Z}\Gamma_\beta^0\right\| = O_p(1)$ by an analogues argument using the $2s$-sparse maximum eigenvalue, and $\left\|\sqrt{\frac{1}{TN}}\sum_t Z_t\Gamma_\beta^0 v_{t+1}^0\right\| = O_p(1)$ by Assumption 4(ii).

Third, $\mathcal{C}_\gamma = o_p(1)$ given

$$\begin{aligned}
\|\mathcal{C}_\gamma\| &= \left\| \sqrt{T}H^\top \left(\frac{\Gamma_\beta^{0\top}\bar{Z}^\top\bar{Z}\Gamma_\beta^0}{N} \right)^{-1} \frac{\Gamma_\beta^{0\top}\bar{Z}^\top}{N}\bar{\epsilon} \right\| \\
&\lesssim_P \left\| \frac{\Gamma_\beta^{0\top}\left(\bar{Z} \pm \mathbb{E}[\bar{Z}]\right)^\top}{N}\sqrt{\frac{1}{T}}\sum_t \epsilon_t \right\| \\
&= \left\| \sqrt{\frac{1}{T}}\sum_t\sum_{j=1}^s \frac{1}{N}\sum_i \Gamma_{\beta,j}^0\left(\bar{Z}_{i,j} - \mathbb{E}[\bar{Z}_{i,j}]\right)^\top \epsilon_{i,t+1} \right\| \\
&\quad + \left\| \sqrt{\frac{1}{T}}\sum_t\sum_{j=1}^s \frac{1}{N}\sum_i \Gamma_{\beta,j}^0\mathbb{E}[\bar{Z}_{i,j}]^\top\epsilon_{i,t+1} \right\| \\
&= O_p(\sqrt{\frac{T}{N}}) = o_p(1)
\end{aligned}$$

where the first equality substitutes the notation; the first probabilistic bound holds by adding and subtracting $\mathbb{E}[\bar{Z}]$ and the previously established results on $\|H\|$ and the spectral norm of the inverse design matrix; the second equality holds by opening up the matrix multiplication for the two terms noting Γ_β^0 selects only s rows of $\bar{Z}^\top$; the final probabilistic bound holds given Assumption 4(i) yields $N^{-1}\sum_i \epsilon_{i,t+1} = O_p(N^{-1/2})$ and a LLN on $T^{-1}\sum_t \bar{Z}_{i,j} - T^{-1}\sum_t \mathbb{E}[\bar{Z}_{i,j}] = o_p(1)$ using the structural moment Assumptions 7; and, the final assumption holds given $\sqrt{T/N} \to 0$.

Thus, given $\mathcal{A}_\gamma + \mathcal{B}_\gamma + \mathcal{C}_\gamma = o_p(1) + O_p(1) + o_p(1) = O_p(1)$, the lemma holds. □

LEMMA 11. *Under the assumptions of Lemma A8, we have*

$$\left\| T^{-1}\hat{V}^\top\hat{V} - T^{-1}V^\top V \right\| = O_p\left(\sqrt{\frac{\log^2(Tp)}{TN^2}}\right).$$

Proof of Lemma A11.

$$\begin{aligned}
\left\| T^{-1}\hat{V}^\top\hat{V} - T^{-1}V^\top V \right\| &= T^{-1}\left\| \hat{V}^\top\hat{V} \pm \widehat{V}^\top V - V^\top V \right\| \\
&\leq T^{-1}\left\| \hat{V}^\top\hat{V} - \widehat{V}^\top V \right\| + T^{-1}\left\| \widehat{V}^\top V - V^\top V \right\| \\
&\leq T^{-1}\left\| \hat{V} \right\| \left\| \hat{V} - V \right\| + T^{-1}\left\| \widehat{V} - V \right\| \|V\| \\
&= O_p\left(\sqrt{\frac{\log^2(Tp)}{TN^2}}\right).
\end{aligned}$$

where the first equality follows by adding and subtracting the term; the first and second inequalities follow by the triangle and Cauchy-Schwartz inequalities, respectively; and, the probabilistic bound follows given $\left\|\widehat{V}\right\| \lesssim_P \sqrt{kT}$ by the normalization, $\left\|\widehat{V} - V^0H\right\| = O_p\left(\frac{\log(Tp)}{N}\right)$ by an analogous argument to Lemma A4; and, $\|V\| \leq \|V^0\|\,\|H\| = O_p(\sqrt{T})O_p(1)$ by Assumption 2(i). □

LEMMA 12. *Under the assumptions of Lemma A8 and Assumption 4(ii), we have*

$$\left\| T^{-1/2}\widehat{V}^\top G - T^{-1/2}V^\top G \right\| = O_p\left(\frac{\log(Tp)}{N}\right).$$

Proof of Lemma A12.

$$\begin{aligned}\left\|T^{-1/2}\widehat{V}^\top G - T^{-1/2}V^\top G\right\| &= T^{-1/2}\left\|\left(\widehat{V}-V\right)^\top\left(V^0HH^{-1}\eta_0+\epsilon^g\right)\right\| \\ &\leq T^{-1/2}\left\|\widehat{V}-V\right\|\left\|V^0\right\|\|\eta_0\| + T^{-1/2}\left\|\widehat{V}-V\right\|\|\epsilon^g\| \\ &\lesssim_P O_p\left(\frac{\log(Tp)}{N}\right).\end{aligned}$$

where the equality follows from definition of the observable factor model; the inequality follows from the use of the triangle and Cauchy-Schwartz inequalities; and, the probabilistic bound follows as in Lemma A11 with $\|\epsilon^g\| \lesssim_P \sqrt{T}$ by Assumption 4(ii). □

LEMMA 13. *Under the assumptions of Lemma A12, we have*

$$\sqrt{T}(\widehat{\eta}-\eta) = O_p(1)$$

Proof of Lemma A13.

$$\begin{aligned}\sqrt{T}(\widehat{\eta}-\eta) &= \sqrt{T}\left(\widehat{\eta}\pm\widetilde{\eta}-\eta\right) \\ &= \left(\frac{\widehat{V}^\top\widehat{V}}{T}\right)^{-1}\sqrt{T}\frac{\widehat{V}^\top G}{T} \pm \left(\frac{V^\top V}{T}\right)^{-1}\sqrt{T}\frac{V^\top G}{T} - \sqrt{T}H^{-1}\eta_0 \\ &= \left(\frac{\widehat{V}^\top\widehat{V}}{T}\right)^{-1}\sqrt{T}\frac{\widehat{V}^\top G}{T} - \left(\frac{V^\top V}{T}\right)^{-1}\sqrt{T}\frac{V^\top G}{T} \\ &\quad + (H^\top)^{-1}\left(\frac{V^{0\top}V^0}{T}\right)^{-1}\sqrt{T}\frac{V^{0\top}\epsilon^g}{T} \\ &= (H^\top)^{-1}\left(\frac{V^{0\top}V^0}{T}\right)^{-1}\sqrt{T}\frac{V^{0\top}\epsilon^g}{T} + o_p(1) = O_p(1).\end{aligned}$$

where the first equality follows from adding and subtracting the infeasible estimator; the second inequality follows from the definitions; the third equality follows from the definition of the model for G and rearranging; the penultimate probabilistic bound follows given the difference between the first two terms is $o_p(1)$ by the results of Lemmas A11 and A12 using an analogous argument as Lemma A10; and, the final probabilistic bound holds given $\|H\| = O_p(1)$ for invertible matrix H, the factors have positive definite second moment matrix given

Assumption 2(ii), the time series mean of the mean zero random variable $v_{t+1}^0 \epsilon_{t+1}^g$ is $\sqrt{T}$ by CLT Assumption 4(iv). □

Define $\Pi_t \coloneqq \sum_{j=1}^{s} \sum_{j'=1}^{s} \Gamma_{\beta,j'} \mathcal{Z}_{t,j,j'} \Gamma_{\beta,j}^\top$ containing nonstochastic scalar $\mathcal{Z}_{t,j,j'}$ from Assumption 4(iv) and the asymptotic variance of the Assumption 4(v) joint CLT as

$$\Phi_{11} = \lim_{p,T,N\to\infty} \frac{1}{T} \mathbb{E}\left[V^\top \epsilon^g \epsilon^{g\top} V\right] \qquad \Phi_{22} = \lim_{p,T,N\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{t'=1}^{T} \mathbb{E}\left[\Pi_t v_{t+1} v_{t'+1}^\top \Pi_{t'}^\top\right]$$

$$\Phi_{12} = \lim_{p,T,N\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{t'=1}^{T} \mathbb{E}\left[v_{t+1} \epsilon_{t+1}^g v_{t'+1}^\top \Pi_{t'}^\top\right]$$

where asymptotic covariance matrix Φ is defined as in the logical way.

Proof of Theorem 1.

$$\begin{aligned}
\sqrt{T}\left(\widehat{\gamma}_g - \gamma_g\right) &= \sqrt{T}\left(\widehat{\eta}^\top \widehat{\gamma} - \eta^\top \gamma\right) \\
&= \left(\widehat{\eta} - \eta\right)^\top \sqrt{T}\left(\widehat{\gamma} - \gamma\right) + \sqrt{T}\eta^\top\left(\widehat{\gamma} - \gamma\right) + \sqrt{T}\gamma^\top\left(\widehat{\eta} - \eta\right) \\
&= \sqrt{T}\eta^\top\left(\widehat{\gamma} - \gamma\right) + \sqrt{T}\gamma^\top\left(\widehat{\eta} - \eta\right) + o_p(1) \\
&= \gamma^{0\top}\left(\frac{V^{0\top}V^0}{T}\right)^{-1} \frac{\sqrt{T}}{T} \sum_t v_{t+1}^0 \epsilon_{t+1}^g \\
&\quad + \eta^{0\top}\left(\frac{\Gamma_\beta^{0\top} \bar{Z}^\top \bar{Z} \Gamma_\beta^0}{N}\right)^{-1} \frac{\Gamma_\beta^{0\top} \bar{Z}^\top}{N} \frac{\sqrt{T}}{T} \sum_t Z_t \Gamma_\beta^0 v_{t+1}^0 \\
&\quad + o_p(1) \\
&= \gamma^{0\top}\left(\frac{V^{0\top}V^0}{T}\right)^{-1} \frac{\sqrt{T}}{T} \sum_t v_{t+1}^0 \epsilon_{t+1}^g + o_p(1) \\
&+ \eta^{0\top}\left(\frac{\Gamma_\beta^{0\top} \bar{Z}^\top \bar{Z} \Gamma_\beta^0}{N}\right)^{-1} \frac{\sqrt{T}}{T} \sum_t \underbrace{\left(\sum_{j=1}^{s} \sum_{j'=1}^{s} \Gamma_{\beta,j'}^0 \frac{1}{N} \sum_i \mathbb{E}[\bar{z}_{i,j'} z_{i,t,j}] \Gamma_{\beta,j}^\top\right)}_{\Pi_t} v_{t+1}^0 \\
&\to_d N\left(0, \sigma_g^2\right)
\end{aligned}$$

where the first equality holds by definition of the estimator and target parameter; the second equality holds by adding and subtracting terms; the third equality holds by Lemma A10

$\sqrt{T}\,(\widehat{\gamma}-\gamma) = O_p(1)$ and Lemma A13 $\widehat{\eta}-\eta = O_P(T^{-1/2}) = o_p(1)$; the last equality holds by Lemmas A10 and A13, which leaves the two non-asymptotically negligible terms at rate $\sqrt{T}$ scaled by the associated true parameters η^0 and γ^0; and, the convergence in distribution holds given by the joint CLT assumption 4(v) applying the delta method.

Define the following two invertible matrices matrices: $A \coloneqq \lim_{T\to\infty} T^{-1}\mathbb{E}\left[V^\top V\right]$ and $B \coloneqq \lim_{p,T,N\to\infty} \frac{1}{N}\mathbb{E}\left[\Gamma_\beta^\top \bar{Z}^\top \bar{Z}\Gamma_\beta\right]$.

The asymptotic variance σ_g^2 is thus given by the delta method:

$$\sigma_g^2 \coloneqq \gamma^\top A^{-1}\Phi_{11}(A^\top)^{-1}\gamma + \eta^\top B^{-1}\Phi_{22}(B^\top)^{-1}\eta + \gamma^\top A^{-1}\Phi_{12}(B^\top)^{-1}\eta + \eta^\top B^{-1}\Phi_{12}^\top(A^\top)^{-1}\gamma.$$

□

1.8.2 Tables

Table 1.1: Monte Carlo Simulations.

p	Parameter	Metric	(1) IPCA	(2) Three-Pass Est.	(3) DSLFM
10	Γ	MSE	0.112526		0.040480
		$Bias^2$	0.020931		0.029007
		Var	0.091596		0.011473
	F	MSE	0.046446	1.023278	1.008919
		$Bias^2$	0.000538	0.006095	0.007407
		Var	0.041890	1.006150	0.992703
	β	MSE	1.736775	0.348060	0.336661
		$Bias^2$	0.051617	0.027838	0.027619
		Var	1.551492	0.008405	0.000433
	C	MSE	0.007724		0.034307
		$Bias^2$	0.000066		0.000184
		Var	0.012636		0.033998
	γ_g	MSE		0.000086	0.000125
		$Bias^2$		0.000003	0.000019
		Var		0.000028	0.000015
		Cov90		0.971000	0.835000
		Cov95		0.990000	0.855000
50	Γ	MSE	0.024564		0.009921
		$Bias^2$	0.008984		0.008385
		Var	0.015580		0.001536
	F	MSE	0.223446	1.034021	1.011574
		$Bias^2$	0.009573	0.033910	0.033418
		Var	0.228714	0.989699	0.967504
	β	MSE	4.171191	0.430072	0.396931
		$Bias^2$	0.606915	0.161588	0.155526
		Var	4.084398	0.013159	0.000983
	C	MSE	0.013972		0.007161
		$Bias^2$	0.000751		0.000212
		Var	0.013849		0.007001
	γ_g	MSE		0.015229	0.014656
		$Bias^2$		0.015084	0.014495
		Var		0.000058	0.000069
		Cov90		1.000000	0.828571
		Cov95		1.000000	0.842857

This table reports Monte Carlo simulations, $S = 200$, for IPCA, Three-Pass Estimators of Giglio and Xiu (2021), and the DSLFM—columns 1, 2, and 3, respectively—for target parameters: latent loadings Γ_β, latent-factors F, average factor loadings $\bar{\beta}$, latent matrix C, and observable factor risk premium γ_g. The true data-generating process has three factors, $N = 500$, $T = 100$, $p \in \{10, 50\}$, and $s = p/10$.

CHAPTER 2

Empirical Crypto Asset Pricing

2.1 Introduction

In this chapter, we investigate the dynamics of crypto asset returns through the lens of factor models. After presenting a set of motivating empirical facts, we provide empirical results to investigate why different crypto assets earn different average returns; conduct inference for crypto's inflation risk premium; and, compare estimation of risk premia of crypto asset excess returns between classic factor models and our new dynamic latent-factor model.

Why Crypto Nakamoto (2008) gifted a novel mechanism design known as Proof-of-Work, enabling a set of adversarial entities to reach consensus on the current state of an open database using cryptography, often framed as a solution to the Byzantine Generals' Problem (Lamport, Shostak, and Pease 1982). The Bitcoin blockchain launched in early 2009, employing Proof-of-Work to pioneer a censorship-resistant digital transaction ledger. This innovation introduced a permissionless payment network for transferring bitcoin, its native digital asset. The emergence of Nakamoto Consensus, along with other blockchain-based consensus mechanisms that followed, enabled the scarcity of digital information, particularly in the form of digital or crypto assets, and thus introduced a new area of economic research.

A New, Attractive, and Independent Asset Class We motivate research into the return dynamics of these crypto assets by establishing the following empirical facts in Section 2.3. The advent of Bitcoin sparked a Cambrian explosion of other crypto assets, evolving

from initial valuations as collectibles into a trillion dollar asset class. Bitcoin has matured into a substantial payment network, settling hundreds of billions of dollars annually, with the large majority of transactions settling for a cost of less than one dollar, thereby offering monetary functions with distributed consensus.

Bitcoin exhibited superior risk-adjusted returns when compared to traditional asset classes over our study period of 2018-2022, inclusive. With regard to independence, bitcoin has lower correlations with the Nasdaq and the S&P500, at 0.23 and 0.21 respectively, as compared to gold's correlations with these indices at 0.26 and 0.28. Moreover, bitcoin's correlation with other assets exhibits significant temporal variance, including several quarters of zero or negative correlation with the Nasdaq; their high correlation (¿ 0.3) is only a recent phenomenon seen in 2022. While these measures are suggestive of an independent asset class, a possibly sufficient statistic is whether there are risk-adjusted return gains from including crypto assets in one's portfolio. From diversifying a risk portfolio of holding 100% Nasdaq to instead holding 60% Nasdaq and 40% the crypto market, one would obtain a Sharpe Ratio gain of 0.53 (from 0.43 to 0.96).

Crypto Signals The emergence of hundreds of crypto assets expands to a new asset class the central focus of empirical asset pricing: the search for explanations of why different assets earn different average returns. A fundamentally unique aspect of the crypto asset class is open state: the state of the digital ledger is readable. This is termed onchain data where one has access to the (onchain) economy's full history of transactions. For instance, we directly observe the hodling time of all Bitcoin wallets to discover a majority of wallets utilize bitcoin as a store of value rather than for speculatively trading.

In this manuscript, we formulate several novel crypto asset characteristics in addition to investigating the signal content of characteristics previously studied in the literature. An additional distinction of this study is to build a panel of tradable crypto asset excess return data with more realistic inclusion criteria than previously studied in the literature.

In examining the signal content across this rich set of asset characteristics, although there are some redundant characteristics and signal decays over the study years, we observe numerous sources of signal for the cross-section of one-week-ahead expected returns. These empirical observations motivate the development of the factor model in Chapter 1, which accommodates time-varying relationships between assets and characteristics, while capable of incorporating and compressing the signal across a large number of characteristics.

Empirical Setting To assemble a weekly panel of tradable crypto assets, we prospectively identify, at the start of each month from 2018 to 2022, inclusive, tradable crypto assets on US centralized crypto exchanges with sufficient trading volume and market capitalization, which results in the number of assets growing from 10 on January 7, 2018 to 204 on December 1, 2022. There are 210 unique assets in the panel.

Motivated by a one percent threshold on an order book's volume, the most restrictive inclusion criteria applied each month, among several other criteria, is for each asset to have a median weekly volume across US exchanges of $500k over the trailing three months. Using this strict set of inclusion criteria to study tradable assets without bias (e.g., price impact on low liquidity crypto assets), our panel has a challenge wherein assets repeatedly enter and leave the panel over time as they rise above and fall below the inclusion criteria. We thus have to reform the panel monthly when fitting models. For example, an asset may not be included in January 2019 even though we have data for it, but then the asset could be included in February 2019, for which, we would want to use all of its historical data (i.e. January 2018-January 2019) to inform our models.

This panel of weekly crypto asset excess returns is not only novel to the literature based on the inclusion criteria but also given it contains several novel asset characteristics across the sixty three characteristics studied within the following categories: onchain, social, financial, momentum, exchange, and microstructure. We do note that five years of data is limited, but this is the current state for empirical crypto asset pricing.

Empirical Applications: Factor Models with Low-Dimensional Characteristics We begin our study of estimating risk premia of crypto asset excess returns by employing classic factor models that use a low-dimensional number of asset characteristics. First, in studying sixty three univariate factors formed as the long-short quintile portfolios sorted on each asset characteristic, we find six financial factors (e.g., three momentums, beta, idiosyncratic skewness, and 5% shortfall) are the only ones associated with significant differences in average one-week-ahead excess returns. Given the these six factors are all functions of strictly previous returns, this is suggestive that crypto asset returns are not driven by fundamental factors.

Next, to build a set of benchmarks for later results, we compare the out-of-sample predictive power in the Q3-Q4 2022 data of three models, namely, a three-factor model of size, crypto market, and momentum; a latent three-factor model fit with PCA; and a dynamic latent-factor model fit with IPCA using a subset of the characteristics. We find IPCA outperforms the other models, suggestive of the signal in the characteristics. Its predictive pricing signal outperforms a random walk and it provides economically and statistically significant risk-adjusted returns for the zero-investment portfolio, whereas the other models underperform a random walk yet provide modest long-short risk-adjusted returns.

Empirical Applications: Factor Models with High-Dimensional Characteristics Utilizing the broader set of asset characteristics, we first establish the comparable out-of-sample predictive ability of the DSLFM compared to the benchmark methods, with supporting bootstrapped characteristic importance measures to elucidate the drivers of returns. Exchanges inflows and outflows were significant characteristics, showing the importance of these onchain measures. While DSLFM achieved a maximum, with one latent factor, out-of-sample Sharpe of 3.3, this underperforms ICPA's maximum, with one latent factor, Sharpe ratio of 4.07.

Additionally, we implement our testing procedure to find that the crypto asset class pro-

vides investors a positive inflation risk premium. Early proponents of Bitcoin and other cryptocurrencies framed these as an outside options or hedges against traditional fiat currencies. To study this question, we use our extended model to recover the 10-year expected inflation mimicking portfolio and measure its risk premium. This inflation risk premium was estimated at a statistically significant 1.4 basis points (0.0097 standard error). This translates to a 7.3% annual excess return, suggestive of positive compensation for investors holding an inflation-hedged crypto portfolio, ceteris paribus.

Relevant Literature For the nascent empirical crypto asset pricing literature, this paper's main contribution is a set of out of sample results of a wide array of factors models in measuring expected returns, from simple univariate to observable multivariate to PCA to IPCA and finally our novel DSLFM. Although a recent literature, there are several dozen papers, of which we will discuss a small selection, studying the performance of all these aforementioned models–which we replicate with a more realistic panel of crypto assets–in addition to comparing to the DSLFM. Liu et al. (2019) uses PCA to price contemporaneous returns and volatility in high-frequency tick data using nine crypto assets. Interestingly, they show how the factor structure changes across market regimes and how the factor model explainability has improved over time. Shams (2020) studies univariate factor performance in a panel of crypto assets, including novel social media and investor base measures. Bianchi and Babiak (2021) shows ICPA outperforms static latent factor models and observable factor models in explaining realized returns as well as measuring expected returns. Using the interpretability of IPCA, the authors note only seven characteristics were significant, including liquidity, momentum, and volatility. Liu, Tsyvinski, and Wu (2022) studies the cross sectional pricing ability of univariate factor portfolio sorts and multivariate observable (i.e., market, size, and momentum) factor model. Several univariate factors have statistically significant results, including characteristics formed using market capitalization, previous prices, momentums, and volumes. The three-factor model accounts for the performance of these

univariate strategies.

Next this paper also contributes novel empirical facts, including motivating crypto as a new asset class, presenting bitcoin's use as a store of value and payment network, and discussing the pricing ability of several new asset characteristics. Makarov and Schoar (2020) provide empirical evidence of arbitrage opportunities between centralized exchanges, in particular across countries as bitcoin often trades at a premium outside the United States. Hu, Parlour, and Rajan (2019) present a set of stylized facts on crypto as a new investable instrument, including crypto's low correlation with gold and equities but significant covariance with bitcoin, which is perhaps driven by the need to purchase bitcoin to access altcoins. Borri (2019) estimate conditional value at risk in a small number of crypto assets to show these crypto assets are highly exposed to tail-risk within the crypto asset class, but not exposed to tail-risk with respect to traditional asset classes. Additionally, the authors present mean variance gains from a small crypto allocation to a traditional portfolio. Bianchi (2020) corroborates these findings showing crypto exhibits a low correlation to traditional asset classes as well as presenting results on previous volumes predicting crypto asset returns. Liu and Tsyvinski (2021) discover momentum and proxies for investor attention explain variation in crypto asset returns, yet major crypto assets do not comove with traditional asset prices nor macroeconomic factors. Zhang and Li (2020), Zhang et al. (2021), and Zhang and Li (2023) present significant risk premia results in a panel of crypto assets for idiosyncratic volatility, downside risk, and liquidity. Bianchi, Guidolin, and Pedio (2022) finds investor attention significantly predicts crypto asset returns. Cheah et al. (2022) finds several asset characteristics and macroeconomic factors with significant predictability for bitcoin returns, while stock and bond market common factors have no correlation with bitcoin returns.

As a final contribution to the empirical crypto asset pricing literature, this paper identifies several improvements for building a realistic panel of tradable crypto assets. For context, there are low barriers to entry in launching a new crypto asset, resulting in tens of thousands of crypto assets with a market as of 2023. However, the vast majority have insufficient liquid-

ity for empirical study given trivial volume would have significant price impact. Moreover, many are not available on a US exchanges. Thus, as we will argue in Section 2.2 that the literature is currently studying asset returns which were not tradable. Our criteria suggests a tradable panel starting with ten crypto assets in 2018 expanding to about two hundred assets at the end of 2022. However, by way of example, in similar time periods, Liebi (2022) studies 652 cryptoassets; Borri et al. (2022) measures risk premia of observable factors using the method of Giglio and Xiu (2021) in a panel of about seven hundred assets; and, Cong et al. (2022) studies risk premiums and factor pricing ability in a panel of four thousand crypto assets.

2.2 Description of Data

Panel Overview We obtain hourly crypto asset prices, trading volumes, and market capitalization from Coin Metrics and CoinAPI for 2018-2022, inclusive.[1] Specifically, we collect these data for all assets with at least four months of trading history and a spot market to USD, USDC, or USDT on one of the following United States exchanges.[2] We exclude years before 2018 given a very small number of relevant assets, and we save 2023 for future out of sample experiments. Price is the volume-weighted average hourly candle mid-price across these exchanges where we average all variables first across the two data providers. [3] Excess returns are formed from this tradable price measure after removing the one month Treasury-bill rate to proxy for the risk-free rate. Our weekly panel is aggregated from this

[1]It should be noted that market capitalization in crypto is difficult to measure given the various measures of token supply. We use Coin Metrics' measure which aims to estimate the free float supply; see https://coinmetrics.io/free-float-supply-a-better-measure-of-market-capitalization/ for further discussion.

[2]Specifically, : Binance US, Bitstamp, Coinbase, Crypto.com, FTX US, Gemini, Kraken, and Kucoin.

[3]To our knowledge, this is a uncommon approach in the literature to use the actually tradable price as opposed to a global price across all exchanges. The mean absolute error between the global price uses the volume-weighted average hourly price and the actual tradable price from the US exchanges is $1.71 while the mean absolute error in weekly return is 95 bps.

hourly panel as we will discuss.

Rolling Inclusion Criteria Crypto assets have a challenging empirical asset pricing problem of how to define a relevant temporarly-changing asset universe. To illustrate, note CoinMarketCap listed the market price of 1381 crypto assets on January 2, 2018. [4] However, the vast majority of these assets are not tradable from the view of empirically studying their returns. That is, as we will discuss in the next paragraph, the tradable volume on US exchanges is only sufficiently high for a small number of assets, such that, historical returns will not be distorted by price impact. Moreover, there is volatility in the trading volume so one must sequentially build an asset universe.

On the first day of each study month, we define an asset as tradable in our study if it has twelve weeks of trailing data to form relevant characteristics; it has a spot market on a US exchange to USD, USDC, or USDT; it is not a stablecoin or synthetic asset (e.g., wrapped BTC, PAXG, etc.); its average market capitalization over the trailing three months was above one basis point of the total crypto market capitalization[5]; it has nonzero trading volume on all trailing twelve weeks; and, finally, its median weekly trading volume is above $500,000 in the trailing twelve weeks.[6]

A nuanced aspect of using this asset universe is to repeatedly build the panel each month to handle assets repeatedly entering and leaving the panel. For example, a previously included asset may not be included in January 2019 so we remove the data we have for it, but then the asset could be included in February 2019, for which, we would want to use its historical data (i.e. January 2018-January 2019) to inform our models. Thus, when fitting all

[4] https://web.archive.org/web/20180102053542/https://coinmarketcap.com/all/views/all/.

[5] Many other papers define a static market capitalization threshold of $1MM, which will not be responsive to the significant expansion and contraction of market capitalizations in the asset class, e.g., Liu, Tsyvinski, and Wu (2022), Liu and Tsyvinski (2021), Shams (2020), Cong et al. (2022), etc.

[6] To our knowledge, this is a novel approach in the empirical crypto asset pricing literature where we use a heuristic of staying below 1% of the volume to assume no price impact and a weekly trade size for a given asset of $5,000. See transaction cost discussion in Section 2.4 for further details.

statistical models in this paper, we have to rebuild the panel at a month by month frequency.

There are a few limitations of this approach. We study only a small number of assets relative to the literature. Moreover, we study only assets available to trade for US investors leaving open to future work a more global view of this asset class. Finally, we identify relevant assets at a monthly frequency for simplicity.

Asset Characteristics For all assets in the universe, we obtain from Q4 2017 through Q4 2022 a rich set of asset characteristics from several data providers.[7] Tables 2.1-2.6 in Appendix 2.6.1 enumerate the details of the characteristics studied across six categories: onchain, exchange, social, momentum, microstructure, and financial. We are replicating results, using our more rigorous panel, for many of these characteristics as derived from the relevant literature, including Bianchi, Guidolin, and Pedio (2022), Borri et al. (2022), Liu, Tsyvinski, and Wu (2022), Liu, Tsyvinski, and Wu (2021), Cong et al. (2022), Liebi (2022), Zhang and Li (2020), Zhang et al. (2021), Zhang and Li (2023), and Yao et al. (2021).

To our knowledge, we also have several asset characteristics novel to the literature across the categories: onchain (i.e., age destroyed, delta flow distribution, delta holders distribution, percentage of supply in profit), exchange (i.e., percentage of supply on various exchanges and exchange flows), social (i.e., positive and negative sentiment, developer activity, and VC owned), momentum (i.e., returns from all time highs and lows), microstructure (i.e., ask and bid sizes), and financial (i.e., market capitalization to realized value).

To handle missing values in the characteristics, we fill with cross-sectional medians, where a characteristic was dropped if it had a week where the majority of assets were missing a value. Although this is restrictive, our aim was to be conservative. Characteristics were normalized on a per model basis as discussed in Section 2.4.

[7]The providers are Coin Metrics, CoinAPI, CoinGecko, CoinMarketCap, Messari, and Santiment.

Panel Statistics To close the description of the data, we present basic statistics for our panel of crypto asset excess returns. Summary statistics are presented in Table 2.7 for the panel's asset returns, market capitalizations, and trading volume. In Panel A, we see the number of assets in the panel begins at 10 in January 2018 and grows to 204 at the end of 2022 with a total of 210 unique assets in the panel. The total market capitalization of our panel captures above 80% of the total crypto market capitalization as reported by CoinMarketCap. The median weekly asset market capitalization is about $840MM, which is decreases over the study period as more small cap assets are included. The median weekly asset trading volume is in the tens of millions of USD across the study time period. In Panel B, we report annualized excess return statistics for the market-weighted return of the crypto assets in the panel (CMKT) at 53.84% per year and a Sharpe ratio of 0.67, which offers a higher absolute and risk-adjusted return than the Nasdaq at 9.85% and 0.43, respectively.

Figure 2.1 reports the empirical distributions of weekly excess returns for the crypto market, bitcoin, and ethereum. The distributions have positive mean with the vast majority of returns between -30% and 30%. Across all three, we see several outlier weekly returns that would be unlikely under a normal DGP. Bitcoin has a tighter distribution than the CMKT. Finally, ethereum interestingly has a right tailed skewness.

Figure 2.2 reports the cumulative excess returns from January 1, 2018 until December 31, 2022 for each asset in the study universe and the crypto market. We thus see 172 of the 210 unique assets ($\sim$82%) have a negative cumulative return over the study period, while 146 and 57 assets ($\sim$70% and $\sim$27%) had a cumulative return below -50% and -90%, respectively.

Finally, Tables 2.8 and 2.9 report descriptive statistics for the panel's dependent variable, asset excess returns over the subsequent week, and the set of sixty three asset characteristics.

2.3 Motivating Empirical Facts

> "We have always had bad money because private enterprise was not permitted to give us a better one...The important truth to keep in mind is that we cannot count on intelligence or understanding but only on sheer self-interest to give us the institutions we need."
>
> —Friedrich A. Hayek (1976) *The Denationalization of Money.*

Tables and Figures 2.3-2.22 present the following eleven motivating empirical facts.

1. From zero in 2009, Bitcoin and hundreds of other crypto assets have become a trillion dollar asset class in 2022, with several multi-billion dollar sub-industries.

2. Bitcoin achieved superior risk-adjusted returns for nearly the entire study time period as compared to traditional asset classes.

3. Bitcoin has lower correlations to the Nasdaq and S&P500 (at 0.23 and 0.21) than that of gold's correlation to these indices (at 0.26 and 0.28).

4. Bitcoin's correlation with other assets is highly time varying, including several quarters of zero or negative correlation with the Nasdaq; their high correlation (> 0.3) is only observed recently in 2022.

5. From diversifying a risk portfolio of holding 100% Nasdaq to 60% Nasdaq and 40% CMKT, one would obtain a Sharpe Ratio gain of 0.53 (from 0.43 to 0.96).

6. The crypto market offers a positive inflation risk premium of 31 bps.

7. Bitcoin is used to store value by a majority of wallets, not speculatively trading.

8. Bitcoin is a payment network settling hundreds of billions of dollars annually where the large majority of transactions cost less then one USD.

9. Efforts to fork, that is copy, the Bitcoin blockchain have had immaterial adverse effects on it; an event study of forks observes, on the contrary, significant positive effects on price, trading volume, active addresses, and social activity.

10. There are several characteristics with significant signal for the cross-section of one-week ahead expected returns.

11. The asset characteristics contain redundant information; however, the variation cannot be captured by just a few principal components.

A Rising Asset Class We begin with documenting the birth and rise of crypto asset class. Figure 2.3 plots the market capitalization, during the study period, of the crypto market, Bitcoin, Ethereum, assets by industry classification, and assets by usage classification. Before the launch of Bitcoin, there were, on the order of, 50 global currencies (Fratzscher 2009). From a market value of zero in 2009, Bitcoin and a Cambrian explosion of hundreds of other cryptocurrencies and crypto assets have risen to a trillion dollar asset class at the end of 2022, with several multi billion dollar sub-industries. The assets included in this study have an aggregate valuation of about $650B as of December 25, 2022, which captures about 80% of the total crypto market, per CoinGecko, as of that date. Bitcoin and Ethereum capture over half of the valuation of this asset class, hence the focus on these assets throughout this paper. In the second two panels of Figure 2.3, we document, using the asset industry and usage classifications of Messari, the rise of several different sub-industries within crypto. We note this not only to emphasize how it is incorrect to conceptualize all crypto assets as cryptocurrencies, but also to note areas for future work to account for these industry and usage differences in asset pricing models and other research.

An Attractive Asset Class Bitcoin achieved superior risk-adjusted returns for nearly the entire study time period as compared to traditional asset classes. Figure 2.4 reports rolling Sharpe ratios over trailing four year windows using weekly excess returns for bitcoin

and various other assets for the study period. To proxy for other assets (i.e., equities, fixed income, real estate, currencies, and gold), various large ETFs were used to capture a low cost exposure to the relevant asset class. Qualitatively, these rolling Sharpe ratios increase in noise as we shrink the window over which we consider an investment horizon; notably, Bitcoin's separation decreases. However, we study this four year window as it is the relevant period over which the median bitcoin is held, as will be discussed shortly.

An Independent Asset Class We now motivate the independence of this asset class by studying correlations and the gains from diversification. Table 2.10 reports pairwise Pearson correlation coefficients between weekly excess returns of bitcoin, ethereum, the crypto market, and various other assets for the January 1, 2018 to December 31, 2022 time period. Bitcoin has lower correlations to the Nasdaq and S&P500 (at 0.23 and 0.21) than that of gold's correlation to these indices (at 0.26 and 0.28). The crypto market is of similar correlations to the Nasdaq and S&P500 (at 0.26 and 0.25) as Gold. This is suggestive evidence of the independence of this asset class; however, this single measure masks significant temporal variation in this relationship between assets.

Table 2.5 reports rolling four-year Pearson Correlations between bitcoin's weekly excess returns and those of other major assets for the study time period. We observe those aggregate positive correlations mask richer temporal variation in this rolling correlation. For example, for twelve months leading up to the COVID 19 onset, Bitcoin had a rolling four-year correlation with Nasdaq of less than 0.1, including a quarter of negative correlation. Correlations between Bitcoin and other assets above 0.3 are only a recent phenomenon in the 2022 calendar year. With risk-adjusted returns and correlation measures as context, we now turn to study a better measure of the independence of this asset class: are there gains from diversifying one's portfolio to include crypto?

Figure 2.6 plots the annualized geometric average return against the annualized volatility

of each crypto asset's weekly excess returns over the study period.[8] Additionally, the risk free rate and the portfolio holding 100% Nasdaq are plotted. Nasdaq, in dark grey, and CMKT, in purple show similar ratios between their annualized geometric average returns and risk with the grey and purple dashed lines stacking nearly on top of each other. However, if we aim to maximize this return-risk ratio in this data set, we should allocate 70% to the Nasdaq and 30% to the crypto market for a ratio of 0.63, which is the portfolio plotted in black. (For the Sharpe Ratio, we'd maximize at 60% Nasdaq and 40% CMKT for a Sharpe of 0.96, a Sharpe gain of 0.53 over 100% Nasdaq at 0.43.) In this data set, the annualized geometric average weekly return of CMKT is 26.4% with an annualized volatility of 80.8%. For BTC, ETH, Nasdaq, and the 70 30 portfolio, those numbers are, respectively: 5.9% and 76.7%; 11.8% and 101.8%; 7.5% and 23.0%; and, 20.5% and 32.4%.

Inflation Risk Premium in the Crypto Asset Class Allocations to bitcoin and the crypto market have been motivated as a hedge on traditional currencies; for example, inflation risk as substantiated by recent survey data (Aiello et al. 2023). Some simple methods to empirically study this claim would be to measure correlations between crypto returns and inflation expectations or, similarly, a price impact study of large changes in inflation expectations on crypto returns. In the spirit of these, if we study only the largest twelve month changes in the Cleveland Fed's 10 year expected inflation measure between January 1 2018 and December 31, 2022, the correlation between bitcoin monthly excess returns and 10 year expected inflation is 0.03 and the correlation between cmkt monthly excess returns and 10 year expected inflation is 0.06, which compared to monthly excess return of Gold and 10 year expected inflation is -0.10. The next level of sophistication would be to partition out the crypto market return in a regression of the monthly excess returns of Bitcoin on the CMKT and expected inflation, which is reported in Panel A of Table 2.11. We observe a positive covariance between inflation expectations and the contemporaneous excess returns

[8]Do note the ordinate axis is a geometric average, not a time series average, given the latter can be quite different from the former given the volatility of these assets (e.g. Zcash has different signs!).

of bitcoin. However, in asset pricing, we have a better tool to capture whether an observable risk factor, e.g., inflation, carries a nonzero risk premium in the crypto asset class.

In Table 2.11, Panel B, we report the positive inflation risk premium of 31 bps per week for the crypto asset class. We utilize Fama Macbeth regressions, which recover the inflation-mimicking portfolio in the crypto asset class and estimate this portfolio's risk premium. Although this yields a positive and economically significant risk premium, it is not statistically distinguishable from zero. Other researchers however have found a similarly positive premium associated with U.S. inflation breakeven rate (Borri et al. 2022). Some further limitations of note: we used a 200 day restriction to calculate the factor loading (Bali, Engle, and Murray 2016); there is potential omitted variable bias of other relevant observable factors (Giglio and Xiu 2021); and, further, a dynamic latent factor model, as opposed to a static observable factor model, will likely better capture the dynamics of returns and regime changes, which we show in Section 2.4.

Onchain Facts on Bitcoin's Use We now turn to study a few onchain statistics, using the rich environment of the Bitcoin blockchain, with the aim to correct a few misconceptions in the literature and to motivate future research in the onchain lab.

First, Bitcoin is used to store value by a majority of wallets, rather than speculatively trading. Figure 2.7 reports the median age in full days of all unspent transaction outputs (UTXO) for the Bitcoin ledger for each week in 2018 through 2022. The majority of UTXOs, which are a proxy for wallets, have not been spent for years. This may be a downward biased estimate given Bitcoin has only been around for about fifteen years, which the trend line supports.

Second, Bitcoin is a payment network settling hundreds of billions of dollars annually where the large majority of transactions cost less then one USD, offering monetary functions with distributed consensus. Figure 2.8 reports Bitcoin's monthly onchain volume and median transaction fee for the study period. We see monthly onchain settled transactions on the

order of tens of billions of USD and median transaction fees paid to miners on the order of one USD.

Third, efforts to fork, that is copy, the Bitcoin blockchain have had immaterial adverse effects on it. Table 2.12 reports this an event study for various Bitcoin statistics on fifteen dates of major Bitcoin forks, subsequent to January 2016.[9] Bitcoin's market value, trading volume, onchain and development activity, social volume, and miner hash rate all respond positively to forks with statistically significant, at the 5% level, positive signs for return, trading volume, active addresses, and social volume. [10] [11]

Asset Characteristic Signal Content Finally, we study some simple statistics to inform the potential signal content of the sixty three asset characteristics in measuring expected returns. Tables 2.13-2.22 report statistics for the panel's asset characteristics, including: univariate regression results of asset excess returns over several forward horizons on each characteristic; the mutual information between asset excess returns seven days ahead and each characteristic for entire study period and each calendar year; and, the correlations among various groups of characteristics and their first principal components. Although decaying over the years, we observe numerous sources of signal for the cross-section of one-week ahead expected returns. Further, the asset characteristics do contain redundant information; however, the variation across asset characteristics cannot be captured by just a few principal components.

[9] We use one week pre and post windows. The fifteen forks considered are: Bitcoin 21, Zcash, Bitcoin Cash, Bitcoin Gold, Bitcoin Diamond, Bitcoin Lightning, Bitcoin Fast, Bitcoin2, BitcoinPlus, Bitcoin Interest, Bitcoin Atom, Bitcoin Private, Microbitcoin, Bitcoin BEP2, and BitcoinSV. These are major assets who had a market price tracked by CMC at some point in time subsequent to the fork. This time period containing fork dates covers both boom and bust markets for BTC.

[10] There are a few major Bitcoin forks not included (e.g. Litecoin), which were outside the time period where we have non-financial data.

[11] To check robustness, we also ran the event study for two day periods before and after event days as well as two week periods before and after. The only qualitative difference in a point estimate were Return and Miner Hash Rate flipping negative for two week windows, albeit both are not statistically indistinguishable from zero. All other statistics maintain sign and significance, or increase significance.

Across the univariate regression results in tables in 2.13 through 2.19, we observe numerous significant coefficients not only contemporaneously and at the horizon to be studied (i.e. seven days ahead), but also that the panel exhibits longer memory with significant results at 14, 30, and 90 day horizons. That is, there is persistent signal in characteristics for returns over multiple future horizons beyond seven days. There are more significant coefficients than we would expect by chance and, moreover, there are numerous coefficients that remain highly significant at all horizons.

In the tables 2.21 and 2.22, we study mutual information to capture a broader, nonlinear measure of the relationship between the asset characteristics and the subsequent seven day asset excess return. Correlations and univariate regression coefficients capture only a simple linear covariance. In these tables, we see similar patterns to the univariate regression results for characteristics with high signal; however, we also observe falling magnitudes in the mutual information.

Across the correlation tables in 2.13 through 2.18, we observe the redundancy in information with many pairs of characteristics having near one Pearson correlation; in the 2.20, we see the first principal components for each asset characteristic category has some redundancy, however the majority of the pairwise relationships show low correlation; and, finally, in the characteristic correlation tables, the first principal components can capture large variation only among a few of the characteristics.

Given the aforementioned results, it suggests we will want to use a model that can (i) compress the redundant information while remaining (ii) rich enough to capture the signal across characteristics and (iii) temporal dependence. Moreover, we will want a model that (iv) allows for time varying relationships between assets and characteristics. We use these empirical facts to motivate the DSLFM.

2.4 Empirical Applications

In this section, we first explore the ability of individual asset characteristics in predicting the cross section of the crypto asset returns. Second, we establish the pricing ability of a variety of benchmark factor models and compare these results to those of the DSLFM. With the DSLFM, we utilize the asset pricing tests developed in Section 1.6 to elucidate the drivers of returns and conduct inference for crypto's inflation risk premium.

Univariate Observable Factor Models We begin our empirical study of the cross-section of returns with a classic nonparametric method of univariate factor portfolio sorts based on the rich set of crypto asset characteristics. We establish a set of empirical patterns in the dynamic of crypto returns to offer suggestive evidence of why different crypto assets earn different average returns. Moreover, these results can be used to motivate and develop theoretical models, in addition to more practical and obvious use.

We study portfolios formed using sixty three asset characteristics as precisely defined in Appendix 2.6.1. The large majority of these have extensively-studied counterparts in equity markets, and further, as previously discussed in Section 2.2, have a small but growing literature studying these risk factors in the crypto asset class. Nevertheless, to our knowledge, we have at least fifteen novel asset characteristics across the six categories we use to group the sixty three characteristics. They are: age destroyed, delta flow distribution, delta holders distribution, percentage of supply in profit, percentage of supply on various exchanges, exchange flows, positive and negative sentiment, developer activity, VC owned, returns from all time highs and lows, ask and bid sizes, and market capitalization to realized value.

To form univariate factor portfolios for the study period 2018-2022, inclusive, we sort crypto assets each week into five value-weighted portfolios ranked by the smallest (portfolio 1) values for the characteristic to the largest (portfolio 5). We analyze the zero-investment portfolio of long the top quintile and short the bottom quintile. That is, each week we sort

individual crypto assets into quintile portfolios based on the value of the given characteristic, and then we track the value-weighted excess return of each portfolio over the week.

Table 2.23 presents the results for the statistically significant zero-investment long-short strategies across all characteristics. Six of sixty three characteristics exhibited statistically significant long-short strategies. Although is a low fraction, this has a low probability of occurring by chance.[12] Across the thirty quintile portfolios, twenty four followed a monotonic pattern within characteristic, with only the two week momentum having monotonic quintile portfolios. The time-series average weekly excess return spread (and annualized Sharpe ratios) for the zero-investment long-short strategies are 1.5% (0.78) for two week momentum, 1.2% (0.75) for one month industry momentum, 1.4% (0.87) for two month industry momentum, 1.6% (0.76) for one week beta, 1.2% (0.79) for one month idiosyncratic skewness, and 1.4% (0.76) for one week 5% expected shortfall.

It is of note that all of these statistically significant strategies were formed on functions of previous returns, that is, momentum and financial characteristics. Although capturing differences in industries lead to significant long momentum strategies, more fundamental pricing characteristics were not associated with significant spreads between top and bottom quintile portfolios. With this more rigorous panel, many of the univariate factor results established in the existing literature fail to replicate. These results motivate the use of the DSLFM to incorporate the signal across these six characteristics with feature selection, in addition to capturing temporal heterogeneity in the relevance of the characteristics through dynamic loadings. Moreover, as we will now see, many characteristics have economically significant return spreads and Sharpe ratios, although the short panel of five years and the volatility of returns in this asset class suppress the resulting t-statistics.

Tables 2.24-2.29 present the results for the statistically insignificant zero-investment long-short strategies across the remaining characteristics, grouped into separate tables by char-

[12]There is a 3.7% probability of observing 6 or more successes out of 63 independent Bernoulli trials with success probability 0.05%.

acteristic category. Five more of these characteristics had an economically meaningful annualized Sharpe ratio above 0.5 for the associated zero-investment long-short strategy. For these, the time-series average weekly excess return spread (and annualized Sharpe ratios) for the zero-investment long-short strategies are 0.8% (0.61) for age destroyed, -1.0% (0.58) for delta holders distribution, -1.0% (0.57) for three to two month reversal, 1.0% (0.56) for one month alpha, and -1.1% (0.54) for one month beta. Seventeen more of the characteristics had an economically meaningful annualized excess return above 30% for the zero-investment long-short strategy.[13] For these, the time-series average weekly excess return spread for the zero-investment long-short strategies are 0.6% for one week transaction volume, 0.7% for percent of circulating supply on a CEX, 0.7% for percent of circulating supply on DeFi, 0.7% for Reddit social volume, 0.8% for one week momentum, 0.6% for one month momentum, -0.6% for return from all time high, -1.0% for return from all time low, 0.9% for one week alpha, 0.7% for one month downside beta, 0.7% for one month coskewness, 0.9% for one week Var5%, -0.6% for one week volatility, -0.6% for one month volatility, -0.9% for three month volatility, -0.5% for one week idiosyncratic volatility, and -0.6% for one month idiosyncratic volatility.

We observe several notable patterns. There is a large Sharpe of economically significant return spreads and Sharpe ratios for the long-short strategies formed using financial characteristics, yet none for microstructure characteristics. Although using these more relaxed economically significant thresholds, there were several onchain characteristics of note, which is promising for more fundamental-based asset pricing. Only a single social characteristic had an economically significant return spread, with no stronger results. These results further motivate the specification of the dynamic latent factor model to select characteristics with signal content for a dynamic factor loading and to compress the common variation into a low-dimensional common risk factor vector.

[13]These are pure-alpha strategies, although for context the CMKT and Nasdaq returned 0.53% and 10% annually during the same time period.

We close this section noting limitations. These univariate factor results do not account for feasibility of short selling and transaction costs, such as spreads, trading fees, margin fees, price impact, and slippage. We also studied results for the entire study period, which includes the out of sample period that we will use to judge the performance of the coming more sophisticated factor models. It is for this reason we did not report annual univariate results, as our out of sample period will comprise only 10%-20% of the total study period.

Multivariate Observable Factor Models We now turn to the out of sample performance of low dimensional factor models in estimating risk premia of crypto asset excess returns, beginning with multivariate observable factor models. Instead of selecting a small number of observable factors based on individual univariate performance, we instead form, using the sixty three characteristics, all combinations of one-, two-, and three-factor models to select the best model of each size based on its performance in combined period of the second half of 2021 and first half of 2022.

In detail, to select multivariate factor models, we perform the following procedure. First, we form strictly time varying risk factors using each of the sixty three asset characteristics as the top minus bottom value-weighted quintile portfolio excess returns. We thus do not normalize characteristics. To form predicted asset returns, we next estimate each asset's static factor loading as the contemporaneous in sample time series regression of excess returns on factor(s); estimate risk factor conditional means as the time series average of the factor in sample; and, predict the asset's return for the next week using the dot product. We use this procedure to fit models in 2018 through the first half of 2021 and, with an expanding window, predict week by week for the combined validation period of the second half of 2021 and the first half of 2022. For all one-factor models, the sixty three choose two two-factor models, and the sixty three choose three three-factor models, we then select the best model of each size based on the predictive R^2 for the fifty two weeks in the validation period. Throughout, we have to reform the panel each month for the relevant included assets. To

compare to the literature, we also formed the Liu, Tsyvinski, and Wu (2022) Fama-French style three-factor model using their CMKT, CSMB, and CMOM risk factors.

Table 2.30 presents results for the test period, i.e., the second half of 2022. The best multivariate observable factor models were size; illiquidity and size; and, size, one month momentum, and three month volatility. Interestingly, the model selection process incorporated size in all three models. The predictive R^2 for all three models and the benchmark model were all negative, performing worse in MSE pricing ability than a random walk. However, although statistically insignificant, all three multivariate observable factor models had economically significant weekly time series average excess return spreads of about 1% with associated Sharpe ratios of 1.31-1.72, which all beat the benchmark model with a return spread of 0.5% and a Sharpe ratio of 0.65. All four associated alphas show these long-short strategy returns are meaningfully uncorrelated. The illiquidity and size two-factor model achieved the highest Sharpe of 1.72, which is suggestive evidence of a low number of factors being optimal. Although the strategies replicated out of sample, we should note these results are again before transaction costs and are for a very small test period.

Static Latent-Factor Model We next study the out of sample performance of a static latent-factor model in estimating risk premia of crypto asset excess returns. We are not only interested in how learning the factors from the data changes the out of sample pricing ability, but also in developing benchmarks for the high-dimensional dynamic latent-factor model. We use the classic approach of PCA (Bai 2003) to estimate one- through five-latent factor models.

In detail, for each month in the test period (i.e., the second half of 2022), we form factor(s) using the matrix of contemporaneous excess asset returns for the relevant included assets; for each asset, we run a time series predictive regression of its weekly excess returns on the factor(s) to obtain its factor loading(s); and, we use each asset's factor loadings and the PCA-estimated factors to generate predicted returns for the out of sample month.

Table 2.30 reports results for the test period, i.e., the second half of 2022. The predictive R^2 for all five latent-factor models were all negative, performing worse in MSE pricing ability than a random walk. Although all five models had positive return spreads, only three of the five models had Sharpe ratios—0.78, 1.24, and 1.34—in the range of the observable factor models; however, there was not a clear pattern across the number of latent factors. This is suggestive evidence of latent factors not offering a clear benefit over the multivariate observable factor models. Their out-of-sample performance yielding uniformly positive Sharpe, generated without using asset characteristics, maintains a benchmark for the richer models.

Dynamic Latent-Factor Model with Low-Dimensional Characteristics We close our study of the out of sample performance of low dimensional factor models in estimating risk premia of crypto asset excess returns. We investigate the performance of IPCA, a dynamic latent-factor model where the number of asset characteristics must be smaller than the number of assets and time periods.[14] Given the small number of assets in the panel (i.e., there are less than two dozen for the majority of the weeks), we have to, outside of IPCA, select features from the sixty three asset characteristics. We chose to just use the characteristics listed in Table 2.23.[15] We again reform the panel each month and normalize period-by-period features to linearly spaced on $[0, 1]$. Finally, we do not specify a constant to allow for mispricing effects but rather explain variation in expected returns using exposure to common latent risk factors.

Table 2.30 reports results for the test period, i.e., the second half of 2022. Predictive R^2 are positive except for a five factor specification with the maximum predictive R^2 of 0.18% for the three-factor model. All five models have economically significant weekly excess return

[14]We are grateful to Matthias Buechner and Leland Bybee for the IPCA implementation at https://github.com/bkelly-lab/ipca.

[15]We realized after doing this that this biases IPCA favorably as the univariate results used the IPCA test period, which is one of several reasons that we have not even formed 2023 data to repeat our out-of-sample exercises in fresh and larger data.

spreads, from 1.5% to 3.1%, and associated annualized Sharpe ratios, from 2.07 to 4.07. The one- through four-factor models have statistically significant time-series average weekly excess return spreads for the zero-investment long-short strategies of, respectively, 2.8%, 2.9%, 3.1%, and 2.4%. The alphas remain statistically and economically significant with little return lost to the market. Remarkably, there are only two quintile portfolios out of twenty five that break monotonicity. Sharpe ratios nearly monotonically decline with the number of factors; the three factor model edges out the two factor model by a difference of 0.13. Although again these results do not account for transaction costs and the test period is short, the dynamic latent-factor model estimated with IPCA dominates the static factor models.

Alpha Tests Before exploring potential gains from incorporating the full set of characteristics into the model, we first investigate whether these factor models can span the six cross-sectional crypto asset return predictors identified in Table 2.23. We study intercepts and loadings in contemporaneous time series regressions, for the whole study time period, of each statistically significant 5-1 univariate strategy on the 5-1 returns for the best—based on Sharpe ratio—multivariate, PCA, and IPCA models.

Table 2.31 reports results. The factor model returns subsume all of the univariate strategies besides one week expected shortfall 5%, which maintains a statistically significant alpha intercept of 1.4% weekly excess return. Half of the strategies have a significant loading on the multivariate risk factor. No strategy has a significant loading on the latent factor models estimated with PCA and IPCA. R^2's are modest from 7% to 22%. Although only one strategy maintained a significant alpha, five of the six strategies maintained economically significant return spreads above 20% annually, which offers suggestive evidence of the possibility of a factor model better spanning these univariate strategies. We turn to the DSLFM with this aim.

Dynamic Latent-Factor Model with High-Dimensional Characteristics We study three questions using the DSLFM. We first compare the out of sample predictability in the same test period to that of the previous factor models, in addition to understanding the characteristics driving returns and to estimating the inflation risk premium in the crypto asset class. The setting is the same weekly panel, reformed each month with the relevant assets, with asset characteristics normalized to 0 to 1.

The DSLFM estimation procedure is outlined in Section 1.4 and the test procedures are defined in Section 1.6. There are several hyperparameters for the statistician to chose, for which we are empirically motivated to use cross validation. Specifically, we generate predicted returns week by week in the same validation period of the second half of 2021 through the first half of 2022 by using an expanding window training data set using all previous weeks from the start of the panel. For models with one to five latent-factors, we cross validate the relevant hyperparameters, including the soft thresholding hyperparameter, the lasso penalty parameter, and the number of trailing weeks to average fitted latent factors over to form predicted factors. We present results for models with the best predictive R^2 in the validation period.

Table 2.32 presents out of sample test period results for the DSLFM. Only one out of the five models had a positive predictive R^2. Eight out of ten models, when forming equal-weighted and value-weighted portfolios within quintile, had economically significant time series average returns for the long-short strategies, which were maintained when studying the associated alphas. The equal-weighted portfolios had in all but one case, the five latent-factor specification, superior Sharpe ratios, driven by both improved return spreads and lower volatility.

Given the small panel, the poor pricing ability of the DSLFM with more factors is perhaps not surprising given it could be over parameterized and noisily estimated. The factor loading matrix grows by p with each additional specified latent factor. Nevertheless, the large majority of the long-short strategies obtained economically significant Sharpe ratios

and associated alphas, representing an improvement over the observable factor models and PCA. However, IPCA outperformed, which is perhaps driven by a meaningful signal to noise ratio improvement through the feature selection done before fitting IPCA. We will explore this in 2023 data, which will yield much more data given the wide cross-section relative to preceding years.

Table 2.33 presents bootstrapped results on asset characteristic importance to understand the drivers of returns. Exchange inflows and outflows were the two statistically significant characteristics with point estimates on their importance more than an order of magnitude larger than the next characteristic. Again, we observe the importance of onchain data. This empirically supports approximate sparsity as a reasonable assumption, given the fast decay with a long tail on the importance of these characteristics. Our theory, although it would accommodate approximate, assumes exact sparsity for simplicity. Interestingly, none of the statistically significant univariate factor strategies were significant in the DSLFM characteristic importance. However, all six were at least in the top half, and, in practice, the importance of studying exchange flows is well known. In future work, we will compare these results to the importance measures available in IPCA.

Finally, to demonstrate the extensibility of the DSLFM, we conduct inference on an observable factor risk premium, namely testing for a nonzero premium for ten year expected inflation risk within the crypto asset class. This has been a long-standing research question to understand the relationship between crypto's returns and inflation. Early proponents of Bitcoin and other cryptocurrencies framed these as an outside option or hedge against traditional fiat currencies. To study this question, we use our extended model with one factor and the associated estimation procedure—as described in Section 1.4 in how we extend Giglio and Xiu (2021)—to recover the 10-year expected inflation mimicking portfolio and measure its risk premium.

The inflation risk premium was estimated to be a statistically significant 1.4 bps with a standard error of 0.0097 bps. This translates to a 7.3% annual excess return, suggestive

of positive compensation for investors holding an inflation-hedged crypto portfolio, ceteris paribus. The result corroborates similar findings using more simple methods, detailed in Section 2.3, with a dynamic latent factor model with superior pricing ability. One could attribute this to several aspects of the DSLFM, for example, it allows for regime changes with time-varying loadings, it incorporates rich structure with the full asset characteristics, among other reasons. There are limitations however, including the slow asymptotic rate with this inference procedure, as discussed in Section 1.5, which is exacerbated by the small cross-section in our setting. Thus, although significant, we should interpret this result as suggestive and seek replication.

2.5 Conclusion

> "Competition would provide better money than would government. I believe we can do much better than gold ever made possible. Free enterprise, i.e. the institutions that would emerge from a process of competition in providing good money, no doubt would.
>
> Two hundred years ago in *The Wealth of Nations* Adam Smith wrote that
>
> 'to expect, indeed, that the freedom of trade should ever be entirely restored in Great Britain, is as absurd as to expect that an Oceana or Utopia should ever be established in it.'
>
> It took nearly 90 years from the publication of his work in 1776 until Great Britain became the first country to establish complete free trade in 1860...I fear that since 'Keynesian' propaganda has filtered through to the masses, has made inflation respectable and provided agitators with arguments which the professional politicians are unable to refute, the only way to avoid being driven by continuing inflation into a controlled and directed economy, and therefore ultimately in order to save civilisation, will be to deprive governments of their power

over the supply of money. What we need now is a Free Money Movement...

I wish I could advise that we proceed slowly. But the time may be short. What is now urgently required is not the construction of a new system but the prompt removal of all legal obstacles which have for two thousand years blocked the way for an evolution which is bound to throw up beneficial results which we cannot now forsee."

—Friedrich A. Hayek (1976) *The Denationalization of Money.*

2.6 Appendix

2.6.1 Details on Crypto Asset Characteristics

The tables below present details for each category of crypto asset characteristics.

Table 2.1: Crypto Asset Onchain Characteristics

Characteristic	Definition
Tx Volume Tm7	The total transaction volume in native units over the trailing seven days.
Active Addresses Tm7	The number of active address over the trailing seven days.
Δ Log New Addresses Tm14-Tm7	The first difference of the logarithm of new addresses from 14 to 7 days ago.
New Addresses Tm7	The total number of new addresses over the trailing seven days.
Total Addresses	The total number of unique addresses.
Circulation Tm7	The number of unique native units transferred over the trailing seven days.
Age Destroyed	The sum over the trailing week of all native units transferred times the number of days since they were previously transferred.
Δ Flow Distribution Tm7	The ratio between the total native units transferred between various entities identified by Santiment (e.g. cex, dexes, defi platforms, whales, etc.) over the trailing week and the total first absolute differences across all the flow variables over the trailing week.
Δ Holders Distribution Tm7	The same functional form as the change in flow but for the total supply held by wallets with various magnitudes, as identified by Santiment, of the total supply.
% Supply in Profit	The percentage of the total native units which last transferred at a market value below the current market value.

Table 2.2: Crypto Asset Exchange Characteristics

Characteristic	Definition
% Circ. Supply CEX	The percentage of circulating supply in native units in wallets associated with CEXs as identified by Santiment.
% Circ. Supply DEX	The percentage of circulating supply in native units in wallets associated with DEXs as identified by Santiment.
% Circ. Supply Defi	The percentage of circulating supply in native units in wallets associated with DeFi platforms as identified by Santiment.
% Circ. Supply Traders	The percentage of circulating supply in native units in wallets associated with active traders as identified by Santiment.
Exchange Inflow	The total number of native units transferred from wallets not associated with exchanges to wallets that are, over the trailing week, as identified by Santiment.
Exchange Outflow	The total number of native units transferred from wallets associated with exchanges to wallets that are not associated with exchanges, over the trailing week, as identified by Santiment.
Number of Trading Pairs	The number of trading pairs identified by CMC on CEXs.

Table 2.3: Crypto Asset Social Characteristics

Characteristic	Definition
Social Volume	The total number of text documents containing the asset name across Reddit, Twitter, Telegram, and BitcoinTalk over the trailing seven days.
Social Volume Reddit	The total number of text documents containing the asset name on Reddit over the trailing seven days.
Social Volume Twitter	The total number of text documents containing the asset name on Twitter over the trailing seven days.
Sentiment Pos. Reddit	The total sentiment score across all text documents with a positive sentiment on Reddit over the trailing seven days.
Sentiment Pos. Twitter	The total sentiment score across all text documents with a positive sentiment on Twitter over the trailing seven days.
Sentiment Neg. Reddit	The total sentiment score across all text documents with a negative sentiment on Reddit over the trailing seven days.
Sentiment Neg. Twitter	The total sentiment score across all text documents with a negative sentiment on Twitter over the trailing seven days.
Developer Activity	The aggregate number of GitHub actions (e.g. commits, forks, comments, etc.), as identified by CoinGecko, over the trailing seven days.
VC Owned	Whether the asset has been funded by a set of prominent venture capitalists as identified by CoinMarketCap.

Table 2.4: Crypto Asset Momentum Characteristics

Characteristic	Definition
Return Tm7	Momentum over the trailing seven days.
Return Tm14	Momentum over the trailing fourteen days.
Return Tm30	Momentum over the trailing thirty days.
Return Tm30	Momentum over the trailing sixty days.
Return Tm90	Momentum over the trailing ninety days.
Return Tm14-Tm7	Short term reversal: difference in return between trailing fourteen and seven days.
Return Tm30-Tm14	Medium term reversal: difference in return between trailing thirty and fourteen days.
Return Tm90-Tm30	Long term reversal: difference in return between trailing ninety and thirty days.
Return from ATH	The return since the all time high price.
Return from ATL	The return since the all time low price.
Return Industry Tm30	Industry momentum over the trailing thirty days.
Return Industry Tm60	Industry momentum over the trailing sixty days.

Table 2.5: Crypto Asset Microstructure Characteristics

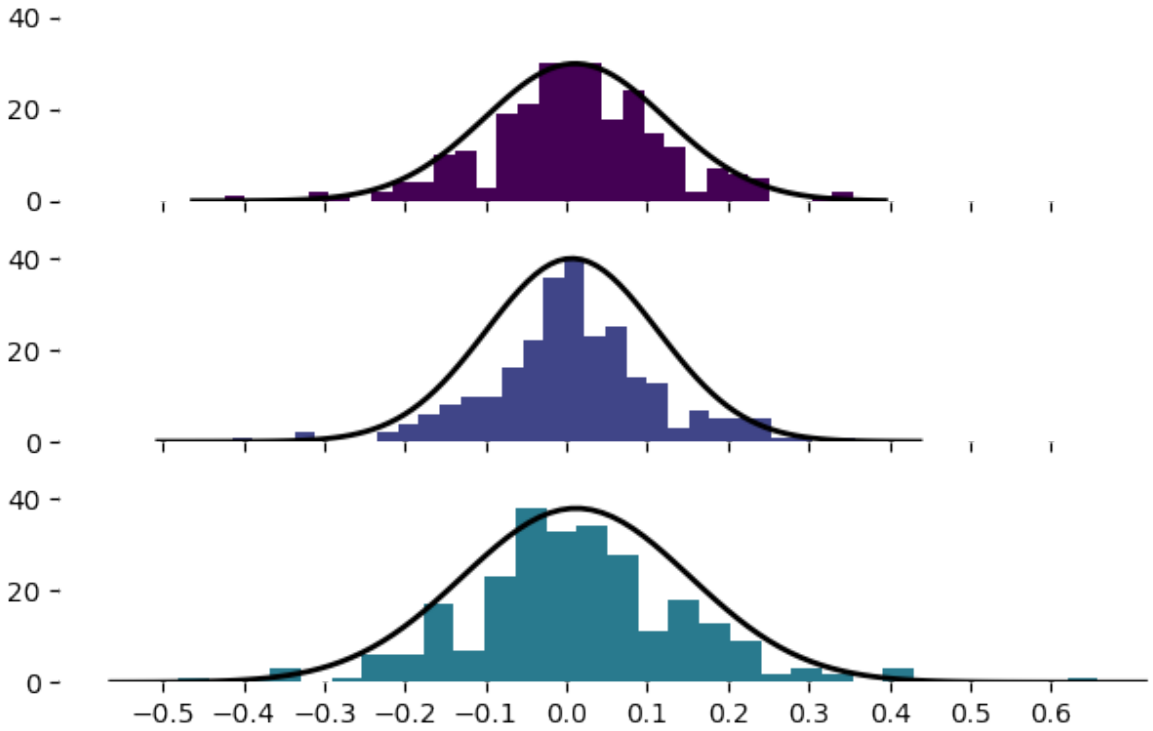
Characteristic	Definition
Trades Sum Tm7	The total number of CEX trades in the trailing seven days.
Volume Sum Tm7	The dollar CEX trading volume in the trailing seven days.
Spread Bps	Spread in basis points.
Ask Size	Market value of orders at best ask.
Bid Size	Market value of orders at best bid.
Illiq Tm7	The average absolute hourly return over the trailing week divided by the average hourly dollar volume over the trailing week.
Turnover Tm7	The total volume over the trailing week divided by the circulating supply in native units.

Table 2.6: Crypto Asset Financial Characteristics

Characteristic	Definition
Price	The market value of one native unit in USD.
Size	The market capitalization of all free floating native units in USD.
MVRV	The ratio of the market capitalization to the realized value, or the total number of free floating native units times the dollar value at the the time of the last transfer.
Alpha Tm7	Intercept coefficient from regressing hourly excess returns on cmkt hourly returns over the trailing seven days.
Alpha Tm30	Intercept coefficient from regressing hourly excess returns on cmkt hourly returns over the trailing thirty days.
Beta Tm7	Slope coefficient from regressing hourly excess returns on cmkt hourly returns over the trailing seven days.
Beta Tm30	Slope coefficient from regressing hourly excess returns on cmkt hourly returns over the trailing thirty days.
Beta Downside Tm30	Slope coefficient from regressing negative hourly excess returns (or zero) on negative cmkt hourly returns over the trailing thirty days.
Coskew Tm30	The slope coefficient on the squared cmkt term from regressing hourly excess returns on cmkt hourly returns and squared cmkt hourly returns over the trailing thirty days.
ISkew Tm30	The skewness of the residuals from from regressing hourly excess returns on cmkt hourly returns and squared cmkt hourly returns over the trailing thirty days.
Shortfall 5% Tm7	Average hourly return of the returns below the fifth quantile of the trailing seven day hourly returns.
VaR 5% Tm7	The fifth quantile of hourly excess returns over the trailing seven days.
Vol Tm7	The standard deviation of hourly excess returns over the trailing seven days.
Vol Tm30	The standard deviation of hourly excess returns over the trailing thirty days.
Vol Tm90	The standard deviation of hourly excess returns over the trailing ninety days.
Ivol Tm7	The standard deviation of the residuals from regressing hourly excess returns on cmkt returns over the trailing seven days.
Ivol Tm30	The standard deviation of the residuals from regressing hourly excess returns on cmkt returns over the trailing thirty days.
Ivol Tm90	The standard deviation of the residuals from regressing hourly excess returns on cmkt returns over the trailing ninety days.

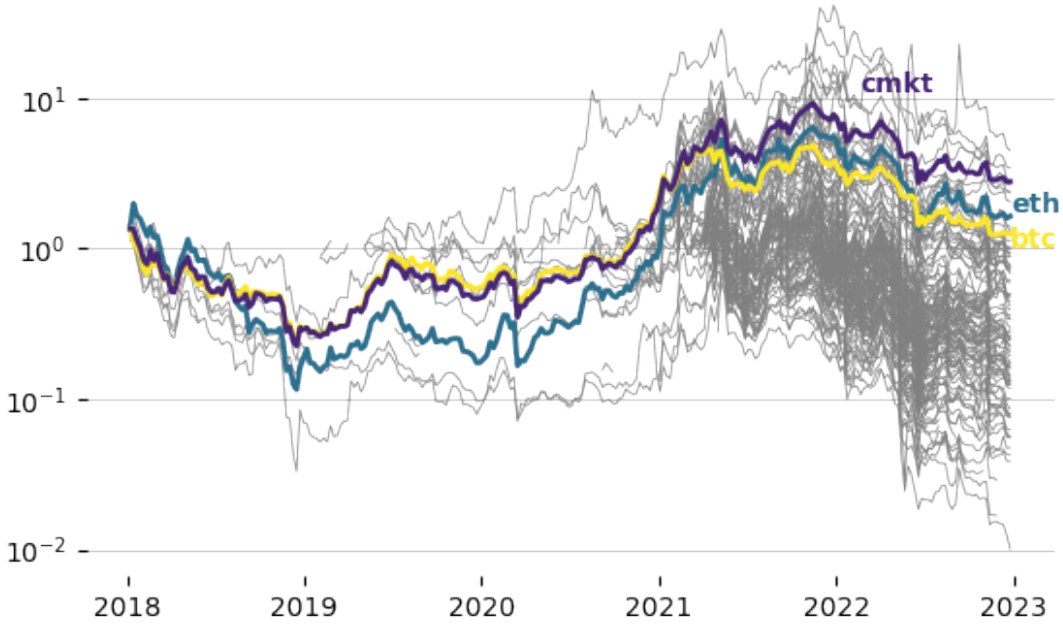
2.6.2 Tables and Figures

Figure 2.1: Empirical distributions of CMKT, Bitcoin, and Ethereum weekly returns.



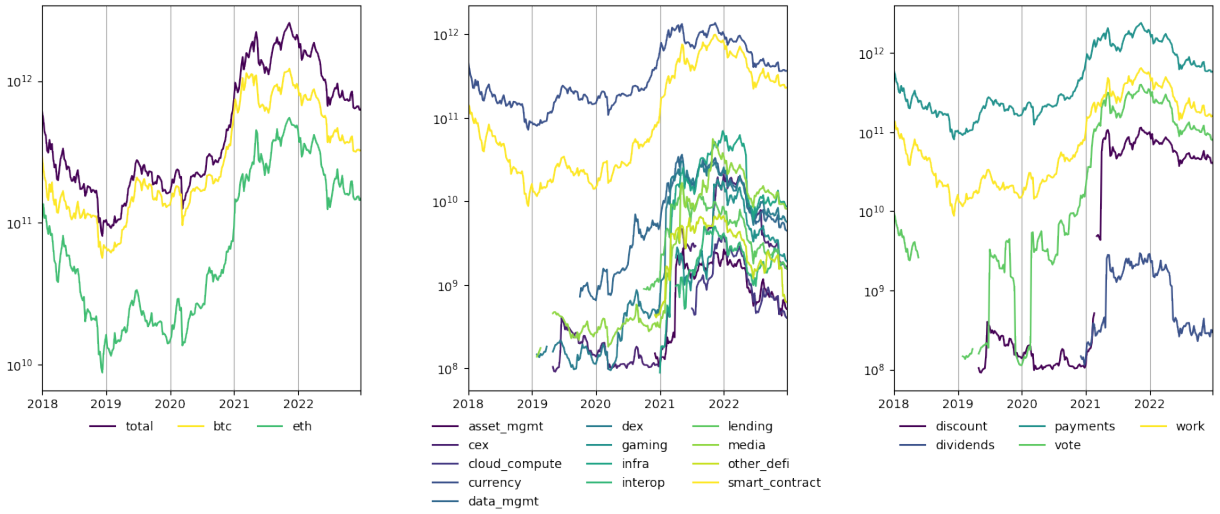
This figure shows the empirical distributions of weekly excess returns, with a normal distribution fit, for coin market (top panel), Bitcoin (middle panel), and Ethereum (bottom panel) for the January 1, 2018 to December 31, 2022 period.

Figure 2.2: Cumulative weekly return of assets in universe.



This figure shows the cumulative excess returns for each asset in the study's universe, and the crypto market, for the January 1, 2018 to December 31, 2022 time period.

Figure 2.3: Market Caps (USD).



This figure shows the market capitalization of: in the first panel, the entire panel, Bitcoin, and Ethereum; in the second panel, the entire panel by asset industry classification; and, in the third panel, the entire panel by asset usage classification. The asset industry and usage classifications are from Messari.

Table 2.7: Summary statistics.

Panel A. Panel summary by year.

Year	Unique Assets	CMKT Excess Return	Total Mcap ($B)	Median Mcap ($B)	Median Volume ($MM)
2018	10	-71.04%	$102	$8.72	$10.27
2019	15	62.89%	$163	$3.70	$11.96
2020	25	280.61%	$618	$2.05	$11.64
2021	154	332.54%	$2,121	$1.42	$27.36
2022	204	-64.05%	$629	$0.45	$14.78
All	210	179.16%	$629	$0.84	$17.58

Panel B. Summary statistics of annualized excess returns.

	Mean	SD	Sharpe	Skewness	Kurtosis	% > 0
CMKT	53.84%	80.61%	0.67	-0.02	0.02	0.53
Bitcoin	27.09%	75.07%	0.36	-0.02	0.02	0.52
Ethereum	52.97%	100.11%	0.53	0.03	0.04	0.52
Nasdaq	9.85%	22.98%	0.43	-0.03	0.03	0.55

Panel C. Extreme events of weekly CMKT excess returns.

Disasters	Counts	%	Miracles	Counts	%
< -5 %	67	25.77%	> 5 %	86	33.08%
< -10 %	35	13.46%	> 10 %	48	18.46%
< -20 %	8	3.08%	> 20 %	14	5.38%
< -30 %	3	1.15%	> 30 %	3	1.15%

This table shows summary statistics on the weekly panel of excess returns from January 1, 2018 to December 31 2022. Panel A reports, by calendar year and for the whole panel, the number of unique assets, the cumulative excess return of the crypto market, the total market capitalization in the last week in billions of dollars, the median market capitalization in billions of dollars, and the median weekly volume in millions of dollars. Panel B reports—for the crypto market (CMKT), Bitcoin, Ethereum, and the Nasdaq—annualized excess return statistics, including the mean, standard deviation, Sharpe ratio, skewness, kurtosis, and percentage of weekly excess returns that are positive. Panel C reports the percentage of extreme events using the weekly crypto market index excess returns.

Table 2.8: Crypto Asset Characteristics: Descriptive Statistics.

			Percentiles						
	Mean	Std	1	5	25	50	75	95	99
LHS									
Asset Excess Return Tp7	-0.004	0.196	-0.402	-0.270	-0.101	-0.015	0.073	0.273	0.615
Onchain									
Tx Volume Tm7	1.1e+11	2.2e+12	4.1e+03	2.2e+05	1.6e+06	6.0e+06	4.3e+07	1.8e+09	1.7e+10
Active Addresses Tm7	2.8e+05	1.1e+06	8.8	428	1.1e+03	2.3e+03	8.9e+03	1.6e+06	7.1e+06
Δ Log New Addresses Tm14-Tm7	-0.030	0.368	-1.1	-0.526	-0.185	-0.053	0.099	0.544	1.1
New Addresses Tm7	3.3e+05	2.2e+06	0.0e+00	1.5e+03	3.8e+03	9.3e+03	3.0e+04	2.9e+05	1.4e+07
Total Addresses	2.3e+06	8.9e+06	1.1e+03	7.2e+03	2.0e+04	2.9e+04	1.1e+05	1.7e+07	4.4e+07
Circulation Tm7	1.1e+11	2.1e+12	1.1e+03	1.5e+05	3.2e+06	8.9e+06	2.2e+07	1.2e+09	9.7e+09
Age Destroyed	6.4e+10	1.7e+12	184	3.3e+04	1.4e+05	5.9e+05	8.7e+06	5.3e+08	4.9e+09
Δ Flow Distribution	-0.483	22	-100	-17	-1.1	-0.029	0.504	15	100
Δ Holders Distribution	0.011	0.027	6.9e-05	6.6e-04	0.003	0.006	0.012	0.034	0.072
% Supply in Profit	58	27	0.0e+00	4.5	37	61	81	97	100
Exchange									
% Circ. Supply CEX	0.195	0.238	1.2e-04	0.001	0.057	0.117	0.226	1.0	1.0
% Circ. Supply DEX	0.012	0.046	0.0e+00	2.4e-06	2.4e-04	0.001	0.005	0.057	0.249
% Circ. Supply Defi	0.017	0.280	0.0e+00	0.0e+00	3.9e-06	3.0e-05	7.1e-04	0.043	0.114
% Circ. Supply Traders	0.544	0.326	5.8e-04	0.005	0.293	0.564	0.824	1.0	1.0
Exchange Inflow Tm7	6.3e+08	6.9e+09	0.0e+00	3.7e+04	6.3e+05	3.2e+06	1.7e+07	4.2e+08	1.5e+10
Exchange Outflow Tm7	6.4e+08	6.9e+09	0.0e+00	4.0e+04	6.5e+05	3.2e+06	1.6e+07	4.0e+08	1.5e+10
Number of Trading Pairs	378	1.2e+03	0.0e+00	1.0	55	115	247	694	8.3e+03
Social									
Social Volume	2.3e+03	9.6e+03	0.0e+00	7.0	58	235	854	1.2e+04	3.9e+04
Social Volume Reddit	23	76	-99	-82	-39	14	82	154	169
Social Volume Twitter	18	64	-95	-72	-31	9.2	60	140	162
Sentiment Pos. Reddit	1.6	6.3	0.0e+00	0.0e+00	0.029	0.157	0.769	5.9	31
Sentiment Pos. Twitter	8.6	45	0.0e+00	0.005	0.115	0.471	2.0	27	199
Sentiment Neg. Reddit	1.6	6.5	0.0e+00	0.0e+00	0.021	0.114	0.683	6.1	32
Sentiment Neg. Twitter	3.1	20	0.0e+00	0.0e+00	0.018	0.098	0.598	8.2	90
Developer Activity	0.054	0.293	-0.413	-0.302	-0.183	-0.004	0.228	0.663	0.836
VC Owned	0.646	0.478	0.0e+00	0.0e+00	0.0e+00	1.0	1.0	1.0	1.0

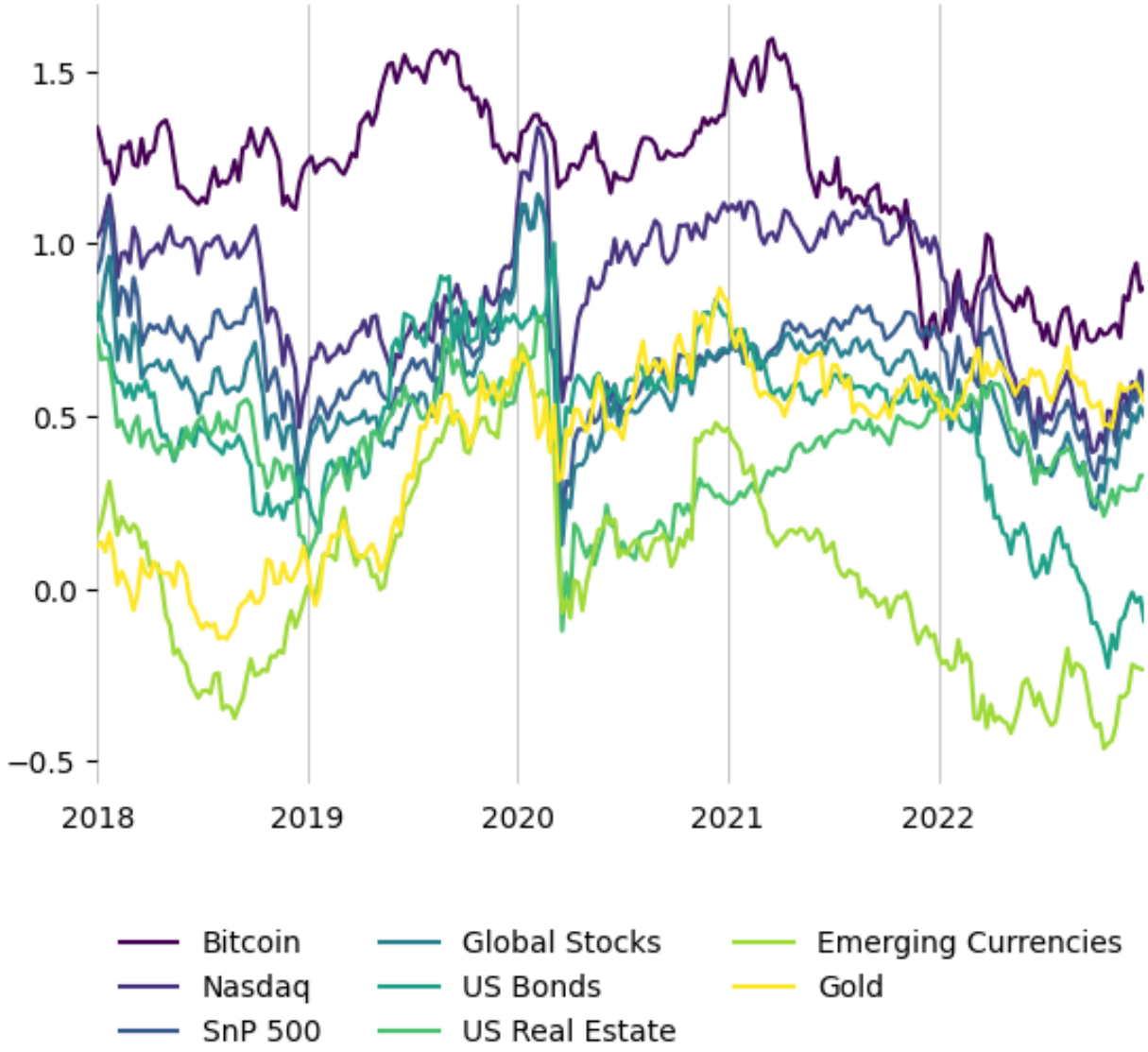
This table reports the summary statistics from the weekly asset panel for the dependent variable, asset excess returns seven days ahead, and the asset characteristics. For each variable, we report the panel mean, median, standard deviation, and selected percentiles. There are 22,678 asset-weeks from January 7, 2018 to December 15, 2022.

Table 2.9: Crypto Asset Characteristics: Descriptive Statistics (Continued).

Momentum									
Return Tm7	0.004	0.906	-0.402	-0.270	-0.101	-0.014	0.075	0.277	0.616
Return Tm14	0.007	1.3	-0.485	-0.348	-0.166	-0.043	0.097	0.409	0.969
Return Tm30	0.010	1.3	-0.599	-0.481	-0.259	-0.095	0.132	0.698	1.7
Return Tm60	0.063	1.5	-0.740	-0.629	-0.385	-0.159	0.174	1.3	3.7
Return Tm90	0.173	2.0	-0.790	-0.676	-0.457	-0.217	0.190	2.2	6.0
Return Tm14-Tm7	0.004	0.907	-0.404	-0.272	-0.104	-0.015	0.074	0.277	0.625
Return Tm30-Tm14	0.016	1.3	-0.522	-0.377	-0.176	-0.039	0.110	0.484	1.2
Return Tm90-Tm30	0.142	1.6	-0.733	-0.620	-0.364	-0.116	0.234	1.6	4.6
Return from ATH	-0.674	0.233	-0.982	-0.948	-0.863	-0.727	-0.540	-0.195	-0.049
Return from ATL	73	1.2e+03	0.014	0.083	0.589	2.8	14	94	370
Return Industry Tm30	0.331	5.4	-0.524	-0.396	-0.187	-0.041	0.204	0.683	1.6
Return Industry Tm60	0.409	5.2	-0.645	-0.552	-0.277	-0.051	0.327	1.4	4.2
Microstructure									
Trades Sum Tm7	2.8e+05	8.5e+05	2.8e+03	7.6e+03	3.3e+04	8.2e+04	2.1e+05	1.1e+06	3.8e+06
Volume Sum Tm7	1.7e+08	7.6e+08	9.1e+05	1.9e+06	5.9e+06	1.8e+07	5.9e+07	5.5e+08	4.4e+09
Spread bps	3.5	28	3.0e-05	2.1e-04	0.001	0.003	0.015	9.3	79
Ask Size	1.6e+04	2.4e+04	11	178	1.2e+03	5.6e+03	2.0e+04	7.3e+04	9.1e+04
Bid Size	3.7e+04	9.7e+04	10	185	1.2e+03	5.6e+03	2.5e+04	2.1e+05	3.5e+05
Illiquidity Tm7	1.7e-07	3.1e-07	1.5e-10	1.9e-09	1.9e-08	6.5e-08	1.9e-07	6.4e-07	1.5e-06
Turnover Tm7	12	130	2.9e-05	3.6e-04	0.005	0.031	0.194	11	327
Financial									
Price	655	4.4e+03	0.003	0.021	0.275	1.8	23	673	2.2e+04
Size	21	2.0	17	18	19	21	22	24	27
MVRV	1.6	3.4	0.0e+00	0.278	0.557	1.1	1.8	4.1	9.2
Alpha Tm7	-2.3e-05	0.003	-0.006	-0.003	-8.4e-04	-2.1e-05	7.2e-04	0.003	0.007
Alpha Tm30	5.7e-05	0.004	-0.001	-6.3e-04	-2.7e-04	-7.7e-05	1.7e-04	7.8e-04	0.002
Beta Tm7	0.010	0.024	-0.040	-0.020	-0.002	0.007	0.020	0.048	0.081
Beta Tm30	0.007	0.017	-0.008	-0.002	0.003	0.006	0.010	0.017	0.025
Downside Beta Tm30	0.021	0.048	-0.018	-0.004	0.007	0.013	0.022	0.045	0.273
Coskew Tm30	-0.001	0.123	-0.220	-0.103	-0.038	-0.005	0.027	0.121	0.240
iSkew Tm30	0.488	1.6	-2.3	-1.1	-0.187	0.215	0.784	2.9	6.8
Shortfall 5% Tm7	-0.024	0.014	-0.077	-0.050	-0.028	-0.020	-0.015	-0.010	-0.007
VaR 5% Tm7	-0.016	0.009	-0.051	-0.035	-0.020	-0.014	-0.010	-0.006	-0.004
Vol Tm7	0.011	0.045	0.003	0.004	0.007	0.009	0.013	0.024	0.036
Vol Tm30	0.013	0.059	0.004	0.005	0.008	0.010	0.014	0.024	0.031
Vol Tm90	0.015	0.064	0.004	0.006	0.009	0.011	0.015	0.024	0.029
Ivol Tm7	0.012	0.069	0.003	0.004	0.007	0.009	0.013	0.024	0.037
Ivol Tm30	0.013	0.074	0.004	0.005	0.008	0.010	0.014	0.024	0.032
Ivol Tm90	0.013	0.064	0.0e+00	0.0e+00	0.008	0.010	0.014	0.023	0.028

This table reports the summary statistics from the weekly asset panel for the dependent variable, asset excess returns seven days ahead, and the asset characteristics. For each variable, we report the panel mean, median, standard deviation, and selected percentiles. There are 22,678 asset-weeks from January 7, 2018 to December 15, 2022.

Figure 2.4: Sharpe Ratios: Bitcoin vs Major Asset Classes.



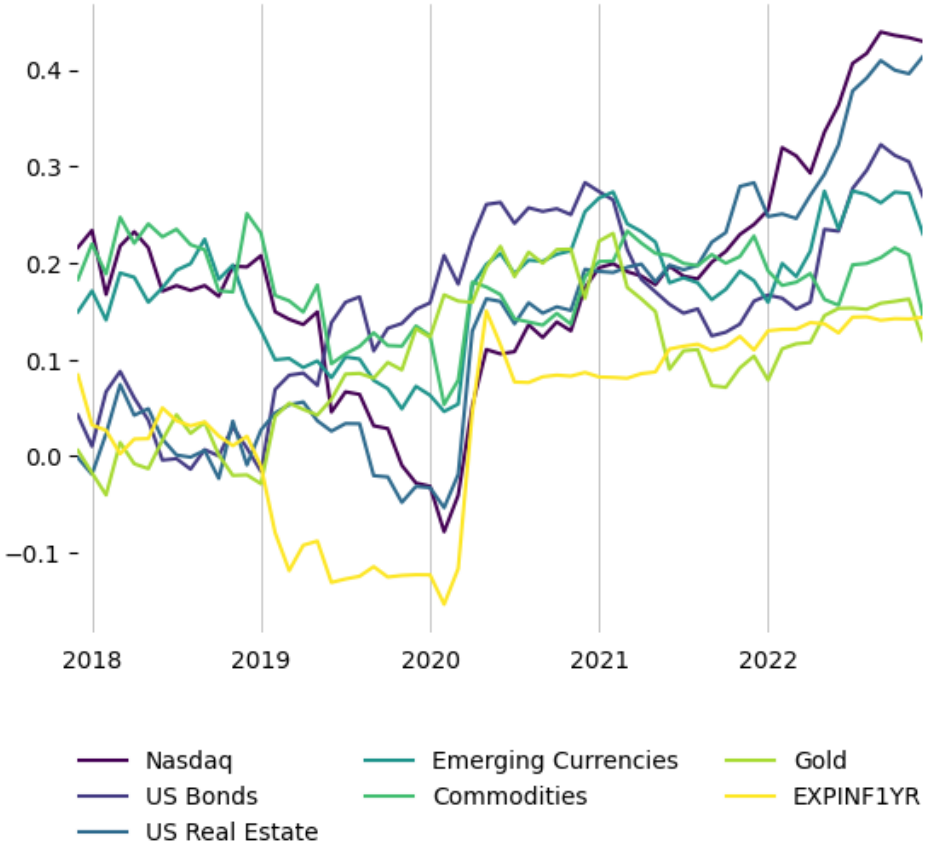
This figure shows the rolling Sharpe Ratio over four year trailing windows using weekly excess returns for various asset classes. Bitcoin is the weekly return from Kraken's order book. Nasdaq and SnP 500 are the returns of the respective indices. The remaining series correspond to the following ETFs: Global Stocks is VT; US Bonds is BND; US Real Estate is VNQ; Emerging Currencies is EBND; and, Gold is GLD.

Table 2.10: Correlations.

		CMKT	BTC	ETH	NSDQ	SP500	RUSS	VT	BND	BNDX	VNQ	EBND	DBC	GLD
Crypto Market	CMKT		0.96	0.89	0.26	0.25	0.28	0.28	0.07	0.04	0.12	0.24	0.24	0.19
Bitcoin	BTC			0.78	0.23	0.21	0.25	0.24	0.06	0.02	0.09	0.21	0.23	0.18
Ethereum	ETH				0.30	0.30	0.31	0.33	0.08	0.07	0.15	0.26	0.24	0.18
Nasdaq	NSDQ					0.95	0.85	0.92	0.30	0.29	0.70	0.57	0.37	0.26
S&P 500	SP500						0.89	0.98	0.31	0.30	0.81	0.62	0.46	0.28
Russell 2000	RUSS							0.92	0.29	0.26	0.79	0.63	0.49	0.28
Global Stocks	VT								0.34	0.30	0.80	0.71	0.49	0.32
US Bonds	BND									0.85	0.48	0.48	0.09	0.47
Ex-US Global Bonds	BNDX										0.43	0.37	0.06	0.36
US Real Estate	VNQ											0.61	0.36	0.33
Emerging Currencies	EBND												0.30	0.46
Commodities	DBC													0.33
Gold	GLD													

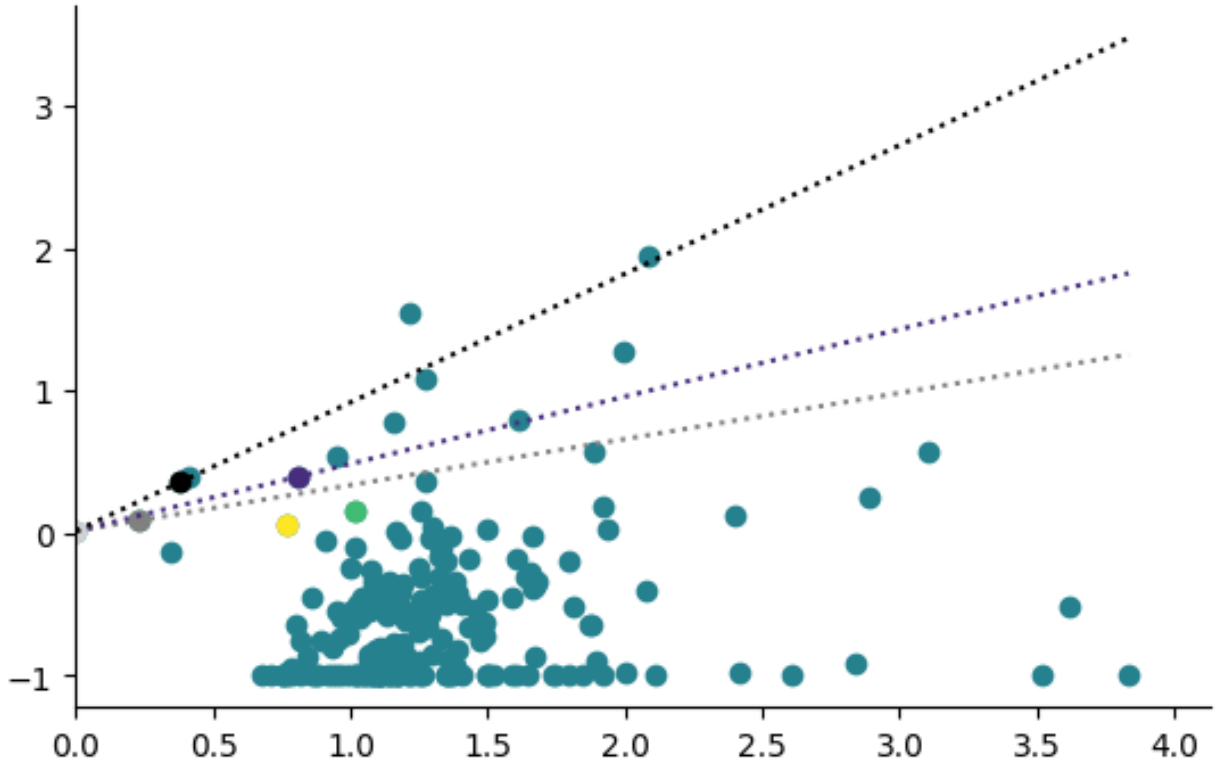
This table reports Pearson correlation coefficients between weekly excess returns of row and column assets for the January 1, 2018 to December 31, 2022 time period. CMKT, Bitcoin, and Ethereum refer to the weekly excess returns of the market cap-weighted assets in the study universe, bitcoin, and ether, respectively. Nasdaq, S&P500, and Russel 2000 refer to the weekly excess return of the IXIC, GSPC, and RUT indices, respectively. Global Stocks, US Bonds, Ex-US Global Bonds, US Real Estate, Emerging Currencies, Commodities, and Gold refer to the weekly excess returns of the following ETFs: VT, BND, BNDX, VNQ, EBND, DBC, and GLD.

Figure 2.5: Rolling Four Year Pearson Correlations: Bitcoin vs Major Asset Classes.



This figure shows rolling four-year Pearson Correlation coefficients between Bitcoin's weekly excess returns and those of other major asset classes for the January 1, 2018 to December 31, 2022 time period. Nasdaq refers to the weekly excess return of the IXIC index. US Bonds, US Real Estate, Emerging Currencies, Commodities, and Gold refer to the weekly excess returns of the following ETFs: BND, VNQ, EBND, DBC, and GLD. EXPINF1YR refers to the Federal Reserve's measure of expected inflation over the subsequent year.

Figure 2.6: Crypto Asset's Annualized Cumulative Returns and Volatility.



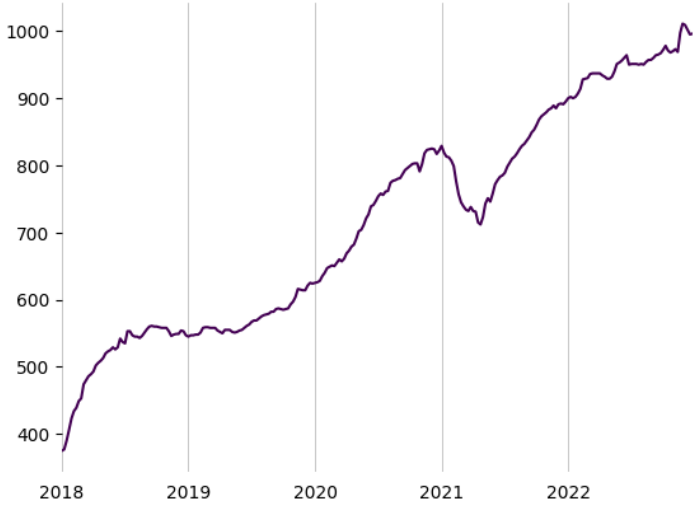
This figure shows the annualized cumulative return and annualized volatility of simple weekly excess returns of the crypto assets in the study universe (over the 2018-2022, inclusive, time period) as well as a few other portfolios. The light grey point is the risk free rate captured by the annualized cumulative return of the 1 month treasury bill during the study period. The grey point is the annualized cumulative return and annualized volatility of the Nasdaq index. For the same two measures, the yellow point corresponds to BTC, the light green point for ETH, and the purple point for CMKT. For the same two measures, the black point corresponds to a portfolio holding 60% Nasdaq and 40% CMKT. The remaining dark green points are for the rest of the assets in the study, removing three assets with outlier returns: DOGE at 9x, LUNA at 52x, and MATIC at 18x.

Table 2.11: Inflation Risk Premium.

Panel A. BTC Return Time-Series Regression.	
Expected Inflation 10 Year	0.1993
	(0.4822)
CMKT	0.3295
	(0.1213)
Constant	0.0180
	(0.0289)
R2	11.7%
N	60
Panel B. Fama-MacBeth Regression.	
Expected Inflation 10 Year	0.0031
	(0.0157)
Constant	0.0373
	(0.0114)
R2	0.2%
N	26

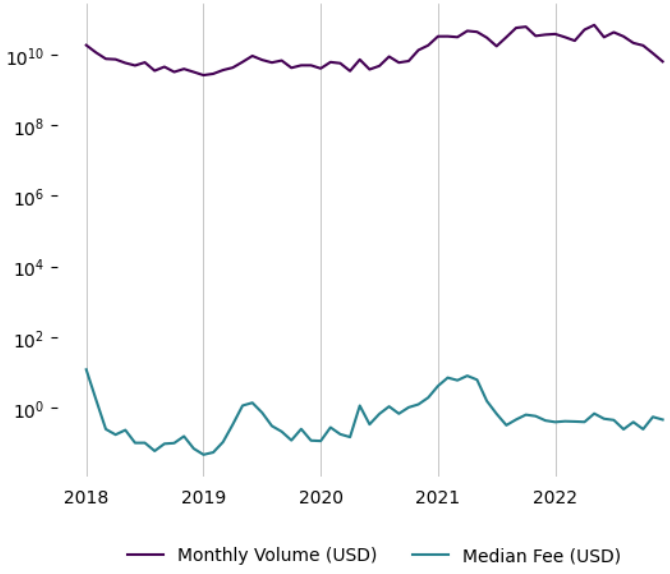
This table reports results from two regressions. Panel A reports point estimates and standard errors from the time-series regression of BTC monthly excess returns on 1 year expected inflation innovations, CMKT monthly excess returns, and a constant. Panel B reports the results from a Fama-MacBeth regression procedure to estimate the risk premium of inflation in the crypto asset class, where we use assets with at least two years of data to precisely estimate beta hats.

Figure 2.7: Hodling: Bitcoin UTXO Median Age in Days.



This figure shows median age in full days of all unspent transaction outputs (UTXOs), rounded down to the nearest day, on the Bitcoin ledger for each week in 2018 to 2022, inclusive. Why hodl?

Figure 2.8: Bitcoin Onchain Transactions.



This figure shows two time series for onchain bitcoin transactions. Monthly Volume reports, in USD, the total calendar month onchain volume transferred between distinct addresses. The Median Fee reports, in USD, the median fee paid to miners across all transactions within each calendar month.

Table 2.12: Bitcoin Forks: Event Study.

	Estimate	Standard Error
Return	0.0079	0.0027
Trading Volume	0.0430	0.0126
Active Addresses	0.0054	0.0029
Developer Activity	0.0174	0.0241
Social Volume	0.0206	0.0061
Miner Hash Rate	0.0001	0.0023

This table reports an event study for various Bitcoin statistics on dates on which fifteen major Bitcoin forks occurred, subsequent to January 2016. The point estimates are the difference between, in the seven days before and after the event date, the average daily change of each characteristic. Return is the daily change in bitcoin's USD price. Trading Volume is the daily change in bitcoin trading volume reported as by CoinMarketCap. Active Addresses is the daily change in the number of unique active Bitcoin addresses as reported by Santiment. Developer Activity is the daily change in the total number of GitHub events (e.g. code pushes, issue interactions, pull requests, comments on commits, etc.) as reported by Santiment. Social Volume is the daily change in the total number of text documents across Reddit, Telegram, Twitter, and BitcoinTalk containing the keyword "bitcoin" as reported by Santiment. Miner Hash Rate is the daily change in the total Bitcoin hash rate as imputed by Coinmetrics. Standard errors are bootstrapped: the standard deviation of the distribution formed by calculating each statistic for 10,000 randomly sampled, with replacement, event days.

Table 2.13: Onchain Characteristics: Correlations and Signal.

Panel A. Correlation of onchain characteristics.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1)	Tx Volume Tm7	1	-0.01	0.04	0.03	-0.01	0.93	0.72	0.01	0.02	0.00
(2)	Active Addresses Tm7		1	0.01	0.40	0.80	-0.01	0.00	0.01	-0.05	0.10
(3)	Δ Log New Addresses Tm14-Tm7			1	0.02	0.02	0.02	0.00	0.14	-0.02	0.00
(4)	New Addresses Tm7				1	0.68	0.02	0.02	0.00	-0.01	0.08
(5)	Total Addresses					1	-0.01	-0.01	0.01	-0.06	0.09
(6)	Circulation Tm7						1	0.87	0.01	0.03	0.00
(7)	Age Destroyed							1	0.00	0.03	0.00
(8)	Δ Flow Distribution								1	0.01	0.02
(9)	Δ Holders Distribution									1	0.08
(10)	% Supply in Profit										1
	PC Onchain	0.93	0.00	0.04	0.04	0.00	0.99	0.91	0.01	0.04	0.00

Panel B. Onchain characteristics exposures.

	0	7	14	30	90
Tx Volume Tm7	6.2e-15*	1.2e-15	3.3e-15	9.9e-15*	2.3e-14*
	(3.7e-15)	(1.4e-15)	(2.4e-15)	(5.2e-15)	(1.3e-14)
	0	0	0	0	0
Active Addresses Tm7	1.0e-09	2.1e-09*	4.8e-10	5.6e-09	1.2e-08**
	(2.6e-09)	(1.2e-09)	(3.7e-09)	(4.3e-09)	(5.6e-09)
	0	0	0	0	0
Δ Log New Addresses Tm14-Tm7	0.09***	0.04***	0.07	0.08**	0.11***
	(0.02)	(7.7e-03)	(0.05)	(0.03)	(0.04)
	0	0	0	0	0
New Addresses Tm7	1.5e-09	1.2e-09*	1.0e-09	3.5e-09**	7.2e-09***
	(9.7e-10)	(6.7e-10)	(1.3e-09)	(1.7e-09)	(2.5e-09)
	0	0	0	0	0
Total Addresses	3.6e-12	3.4e-10*	1.7e-10	1.0e-09	2.1e-09***
	(3.6e-10)	(1.8e-10)	(5.3e-10)	(6.3e-10)	(8.2e-10)
	0	0	0	0	0
Circulation Tm7	4.1e-15	1.7e-15	4.0e-15	1.1e-14	2.2e-14
	(3.8e-15)	(1.8e-15)	(3.3e-15)	(7.5e-15)	(1.4e-14)
	0	0	0	0	0
Age Destoryed	2.2e-15	1.8e-17	5.1e-16	3.7e-15	7.6e-15
	(3.1e-15)	(8.0e-16)	(1.9e-15)	(4.7e-15)	(9.5e-15)
	0	0	0	0	0
Δ Flow Distribution	-4.9e-05	1.6e-04	6.0e-05	5.2e-05	-3.6e-04
	(1.0e-04)	(1.0e-04)	(2.2e-04)	(2.4e-04)	(3.4e-04)
	0	0	0	0	0
Δ Holders Distribution	0.25	0.21***	0.17	0.39	0.51
	(0.20)	(0.06)	(0.12)	(0.24)	(0.39)
	0	0	0	0	0
% Supply in Profit	1.2e-03***	4.6e-04***	3.5e-06	7.9e-04	1.3e-03*
	(4.3e-04)	(1.2e-04)	(5.9e-04)	(6.5e-04)	(7.5e-04)
	0	0	0	0	0
PC Onchain	6.6e-03	1.7e-03	4.2e-03	0.01	0.03
	(5.6e-03)	(2.0e-03)	(3.9e-03)	(8.9e-03)	(0.02)
	0	0	0	0	0

This table reports the correlation matrix among Onchain Characteristics and the loadings on asset excess returns on each characteristic at various horizons. Panel A reports pairwise Pearson correlation coefficients among the characteristics and the first principal component from them. The characteristics are re-scaled to be mean zero and unit variance before PCA and studying these correlations. Panel B reports the coefficient (with 1, 2, and 3 stars for significant at the 10%, 5%, and 1% levels, respectively), standard error, and R^2 for univariate panel regressions of asset excess weekly returns at 0, 7, 14, 30, and 90 days ahead on each of the characteristics and a constant. Standard errors are Newey-West adjusted using Bartlett's formula for the number of lags. There are 22,678 observations.

Table 2.14: Exchange Characteristics: Correlations and Signal

Panel A. Correlation of exchange characteristics.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1)	% Circ. Supply CEX	1	0.10	0.13	0.65	0.26	0.26	0.39
(2)	% Circ. Supply DEX		1	0.01	0.10	0.06	0.06	0.07
(3)	% Circ. Supply Defi			1	0.06	0.00	0.00	0.00
(4)	% Circ. Supply Traders				1	0.11	0.11	0.20
(5)	Exchange Inflow Tm7					1	1.00	0.53
(6)	Exchange Outflow Tm7						1	0.53
(7)	Number of Trading Pairs							1
	PC Exchange	0.62	0.16	0.06	0.45	0.87	0.87	0.75

Panel B. Exchange characteristic exposures.

	0	7	14	30	90
% Circ. Supply CEX	-1.0e-02	0.01	-8.9e-03	0.03	0.09**
	(0.02)	(8.7e-03)	(0.03)	(0.03)	(0.04)
	0	0	0	0	0
% Circ. Supply DEX	-1.3e-02	0.06	0.01	0.13	0.34**
	(0.06)	(0.04)	(0.10)	(0.12)	(0.15)
	0	0	0	0	0
% Circ. Supply Defi	-6.4e-03	-3.1e-03	-1.2e-02	-2.1e-02	-3.2e-02*
	(6.8e-03)	(6.6e-03)	(0.01)	(0.02)	(0.02)
	0	0	0	0	0
% Circ. Supply Traders	-6.0e-03	6.9e-03	-1.4e-02	6.5e-03	0.05*
	(0.01)	(6.7e-03)	(0.02)	(0.02)	(0.03)
	0	0	0	0	0
Exchange Inflow	6.1e-14	9.0e-14	-4.8e-14	1.6e-13	7.3e-13*
	(1.7e-13)	(1.4e-13)	(2.4e-13)	(3.2e-13)	(3.8e-13)
	0	0	0	0	0
Exchange Outflow	6.3e-14	9.3e-14	-4.2e-14	1.7e-13	7.4e-13*
	(1.7e-13)	(1.4e-13)	(2.4e-13)	(3.2e-13)	(3.8e-13)
	0	0	0	0	0
Number of Trading Pairs	6.8e-07	3.0e-06***	2.0e-06	8.3e-06*	1.9e-05***
	(2.5e-06)	(1.0e-06)	(3.8e-06)	(4.3e-06)	(5.5e-06)
	0	0	0	0	0
PC Exchange	-2.6e-04	1.6e-03*	-5.0e-04	3.6e-03	0.01***
	(2.2e-03)	(9.4e-04)	(3.3e-03)	(3.6e-03)	(4.3e-03)
	0	0	0	0	0

This table reports the correlation matrix among Exchange Characteristics and the loadings on asset excess returns on each characteristic at various horizons. Panel A reports pairwise Pearson correlation coefficients among the characteristics and the first principal component from them. The characteristics are re-scaled to be mean zero and unit variance before PCA and studying these correlations. Panel B reports the coefficient (with 1, 2, and 3 stars for significant at the 10%, 5%, and 1% levels, respectively), standard error, and R^2 for univariate panel regressions of asset excess weekly returns at 0, 7, 14, 30, and 90 days ahead on each of the characteristics and a constant. Standard errors are Newey-West adjusted using Bartlett's formula for the number of lags. There are 22,678 observations.

Table 2.15: Social Characteristics: Correlations and Signal.

Panel A. Correlation of social characteristics.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1)	Social Volume	1	0.32	0.33	0.64	0.86	0.66	0.54	0.26	0.05
(2)	Social Volume Reddit		1	0.72	0.32	0.19	0.31	0.13	0.40	-0.01
(3)	Social Volume Twitter			1	0.31	0.25	0.30	0.16	0.41	0.05
(4)	Sentiment Pos. Reddit				1	0.48	0.98	0.67	0.24	0.00
(5)	Sentiment Pos. Twitter					1	0.50	0.61	0.11	-0.01
(6)	Sentiment Neg. Reddit						1	0.70	0.23	0.01
(7)	Sentiment Neg. Twitter							1	0.05	-0.04
(8)	Developer Activity								1	0.30
(9)	VC Owned									1
	PC Social	0.85	0.52	0.54	0.87	0.75	0.88	0.74	0.38	0.05

Panel B. Exchange characteristic exposures.

	0	7	14	30	90
Social Volume	1.1e-07	-1.5e-07	-1.1e-06**	-1.6e-06***	-2.8e-06***
	(4.6e-07)	(1.9e-07)	(5.7e-07)	(6.2e-07)	(7.1e-07)
	0	0	0	0	0
Social Volume Reddit	2.2e-04***	9.9e-05***	2.6e-04***	5.2e-04***	8.0e-04***
	(6.3e-05)	(2.9e-05)	(5.2e-05)	(9.1e-05)	(1.1e-04)
	0	0	0	0	0
Social Volume Twitter	9.7e-05	1.0e-04***	6.6e-05	3.1e-04**	5.5e-04***
	(8.2e-05)	(3.6e-05)	(1.2e-04)	(1.4e-04)	(1.7e-04)
	0	0	0	0	0
Sentiment Pos. Reddit	3.5e-03	1.3e-04	-5.7e-04	1.9e-03	2.0e-03
	(2.2e-03)	(4.2e-04)	(8.6e-04)	(2.1e-03)	(2.2e-03)
	0	0	0	0	0
Sentiment Pos. Twitter	2.6e-04	-8.0e-05	-3.0e-04***	-1.5e-04	-5.7e-04
	(3.4e-04)	(4.9e-05)	(1.1e-04)	(3.4e-04)	(3.5e-04)
	0	0	0	0	0
Sentiment Neg. Reddit	3.1e-03	5.0e-05	-6.9e-04	1.8e-03	1.8e-03
	(2.2e-03)	(4.1e-04)	(7.5e-04)	(2.1e-03)	(2.2e-03)
	0	0	0	0	0
Sentiment Neg. Twitter	2.6e-03	-4.1e-04*	-1.1e-03***	1.7e-03	1.4e-03
	(3.1e-03)	(2.4e-04)	(2.8e-04)	(3.1e-03)	(3.1e-03)
	0	0	0	0	0
Developer Activity	0.05	9.7e-03	0.04***	0.10***	0.15***
	(0.04)	(6.2e-03)	(0.02)	(0.04)	(0.05)
	0	0	0	0	0
VC Owned	-2.6e-02	9.1e-04	-5.7e-02	-5.9e-02	-6.8e-02
	(0.03)	(4.4e-03)	(0.04)	(0.04)	(0.05)
	0	0	0	0	0
PC Social	9.8e-03	3.4e-04	-1.9e-03	7.1e-03	7.3e-03
	(7.1e-03)	(1.1e-03)	(2.2e-03)	(7.2e-03)	(7.4e-03)
	0	0	0	0	0

This table reports the correlation matrix among Social Characteristics and the loadings on asset excess returns on each characteristic at various horizons. See the similar table descriptions for further detail.

Table 2.16: Momentum Characteristics: Correlations and Signal.

Panel A. Correlation of momentum characteristics.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1)	Return Tm7	1	0.71	0.02	0.05	0.05	0.00	-0.01	0.01	0.04	0.53	0.00	0.03
(2)	Return Tm14		1	0.04	0.08	0.07	0.70	-0.01	0.01	0.06	0.87	0.00	0.05
(3)	Return Tm30			1	0.23	0.17	0.02	0.97	0.01	0.14	0.28	0.23	0.15
(4)	Return Tm60				1	0.79	0.05	0.11	0.50	0.54	-0.01	-0.01	0.63
(5)	Return Tm90					1	0.04	0.08	0.75	0.49	0.00	-0.01	0.50
(6)	Return Tm14-Tm7						1	-0.01	0.01	0.04	0.70	-0.01	0.04
(7)	Return Tm30-Tm14							1	0.00	0.06	0.29	0.23	0.08
(8)	Return Tm90-Tm30								1	0.38	0.00	-0.03	0.36
(9)	Return from ATH									1	-0.01	-0.04	0.53
(10)	Return from ATL										1	0.07	-0.01
(11)	Return Industry Tm30											1	0.00
(12)	Return Industry Tm60												1
	PC Momentum	0.26	0.38	0.41	0.82	0.82	0.28	0.32	0.65	0.66	0.36	0.06	0.69

Panel B. Momentum characteristic exposures.

	0	7	14	30	90
Return Tm7	1.00***	2.3e-03**	-1.9e-02	0.97***	0.98***
	(8.2e-16)	(1.1e-03)	(0.02)	(0.03)	(0.02)
	1	0	0	0.44	0.4
Return Tm14	0.50	-9.7e-04	-1.3e-02	0.48	0.49
	(0.34)	(3.1e-03)	(8.2e-03)	(0.34)	(0.34)
	0.5	0	0	0.22	0.2
Return Tm30	0.02	3.2e-03*	-1.8e-03	5.4e-03	4.8e-03
	(0.01)	(1.8e-03)	(6.5e-03)	(7.6e-03)	(8.0e-03)
	0	0	0	0	0
Return Tm60	0.05***	8.3e-03***	-7.9e-03	0.02	0.02
	(0.01)	(3.5e-03)	(0.02)	(0.02)	(0.02)
	0	0	0	0	0
Return Tm90	0.03***	6.3e-03***	3.4e-03	0.01*	7.7e-03
	(5.6e-03)	(2.1e-03)	(6.5e-03)	(7.6e-03)	(8.5e-03)
	0	0	0	0	0
Return Tm14-Tm7	-4.2e-03	-4.9e-03***	-1.6e-02	-1.6e-02	-2.1e-02
	(9.8e-03)	(1.1e-03)	(0.01)	(0.01)	(0.02)
	0	0	0	0	0
Return Tm30-Tm14	-4.8e-03	2.8e-03*	3.8e-03	-4.0e-03	-6.6e-03
	(5.4e-03)	(1.5e-03)	(2.4e-03)	(5.3e-03)	(5.9e-03)
	0	0	0	0	0
Return Tm90-Tm30	0.01***	3.5e-03	0.01	4.7e-03	-7.9e-03
	(3.2e-03)	(2.9e-03)	(6.9e-03)	(8.7e-03)	(9.7e-03)
	0	0	0	0	0
Return from ATH	0.19***	0.03**	-8.4e-02	0.05	0.11
	(0.07)	(0.01)	(0.11)	(0.11)	(0.12)
	0	0	0	0	0
Return from ATL	6.9e-04	-3.9e-06	-1.1e-05***	6.8e-04	6.7e-04
	(6.0e-04)	(4.8e-06)	(1.7e-06)	(6.1e-04)	(6.1e-04)
	0.28	0	0	0.13	0.11
Return Industry Tm30	6.6e-04	8.7e-04**	2.6e-04	1.5e-03***	2.2e-03***
	(5.7e-04)	(4.1e-04)	(4.7e-04)	(5.7e-04)	(8.2e-04)
	0	0	0	0	0
Return Industry Tm60	0.04***	0.01***	-2.7e-03	0.02	0.03*
	(0.01)	(3.7e-03)	(0.01)	(0.02)	(0.02)
	0	0	0	0	0
PC Momentum	0.14	5.0e-03***	-4.4e-03	0.13	0.13
	(0.10)	(2.1e-03)	(9.2e-03)	(0.11)	(0.11)
	0.07	0	0	0.02	0.02

See the similar table descriptions for further detail.

Table 2.17: Microstructure Characteristics: Correlations and Signal.

Panel A. Correlation of microstructure characteristics.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1)	Trades Sum Tm7	1	0.68	-0.03	0.09	0.12	-0.15	0.10
(2)	Volume Sum Tm7		1	-0.03	0.20	0.23	-0.12	0.20
(3)	Spread bps			1	0.15	0.12	0.12	-0.01
(4)	Ask Size				1	0.48	0.11	0.01
(5)	Bid Size					1	0.02	0.03
(6)	Illiquidity Tm7						1	-0.04
(7)	Turnover Tm7							1
	PC Microstructure	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Panel B. Microstructure characteristics exposures.

	0	7	14	30	90
Trades Sum Tm7	1.9e-07	-7.7e-09***	8.1e-07*	4.3e-07	4.1e-07
	(1.8e-07)	(3.0e-09)	(4.5e-07)	(3.2e-07)	(3.2e-07)
	0.02	0	0.22	0.06	0.05
Volume Sum Tm7	2.0e-11**	-2.8e-13	8.3e-11	2.6e-11	2.2e-11
	(9.8e-12)	(1.7e-12)	(6.6e-11)	(2.0e-11)	(2.1e-11)
	0	0	0	0	0
Spread bps	-1.7e-04**	1.3e-04	5.2e-04*	6.2e-04*	3.3e-03***
	(8.1e-05)	(9.1e-05)	(2.7e-04)	(3.4e-04)	(8.5e-04)
	0	0	0	0	0
Ask Size	2.1e-07	2.4e-07*	-8.6e-08	6.7e-07	2.7e-06***
	(2.9e-07)	(1.3e-07)	(4.2e-07)	(5.0e-07)	(6.0e-07)
	0	0	0	0	0
Bid Size	1.9e-07***	7.8e-08**	-3.7e-08	2.1e-07*	6.6e-07***
	(7.8e-08)	(3.5e-08)	(7.3e-08)	(1.1e-07)	(2.3e-07)
	0	0	0	0	0
Illiquidity Tm7	-3.8e+04***	35004.30***	10815.31	112251.74***	550580.38***
	(8346.09)	(7319.20)	(27987.45)	(33168.74)	(58687.65)
	0	0	0	0	0
Turnover Tm7	-1.2e-07	-2.1e-06	-2.0e-05	-4.8e-05***	-8.6e-05***
	(1.6e-05)	(9.0e-06)	(1.4e-05)	(1.9e-05)	(2.2e-05)
	0	0	0	0	0
PC Microstructure	0.05	5.1e-04	0.20	0.11	0.13
	(0.04)	(1.8e-03)	(0.15)	(0.08)	(0.08)
	0.01	0	0.04	0.01	0.01

This table reports the correlation matrix among Microstructure Characteristics and the loadings on asset excess returns on each characteristic at various horizons. Panel A reports pairwise Pearson correlation coefficients among the characteristics and the first principal component from from them. The characteristics are re-scaled to be mean zero and unit variance before PCA and studying these correlations. Panel B reports the coefficient (with 1, 2, and 3 stars for significant at the 10%, 5%, and 1% levels, respectively), standard error, and R^2 for univariate panel regressions of asset excess weekly returns at 0, 7, 14, 30, and 90 days ahead on each of the characteristics and a constant. Standard errors are Newey-West adjusted using Bartlett's formula for the number of lags. There are 22,678 observations.

Table 2.18: Financial Characteristics: Correlations.

Panel A. Correlation of financial characteristics.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
(1)	Price	1	0.32	0.00	0.00	0.00	-0.03	-0.02	-0.01	0.00	-0.04	0.09	0.10	-0.01	-0.01	-0.02	-0.01	-0.01	-0.01
(2)	Size		1	0.01	-0.01	-0.01	-0.03	-0.04	0.00	0.04	-0.17	0.14	0.16	-0.02	-0.03	-0.03	-0.02	-0.02	-0.01
(3)	MVRV			1	-0.01	0.00	-0.01	0.00	0.02	0.01	0.04	-0.06	-0.07	0.01	0.00	0.01	0.00	0.00	-0.01
(4)	Alpha Tm7				1	0.54	0.26	0.42	0.00	0.52	0.16	-0.16	-0.11	0.80	0.37	0.37	0.79	0.37	0.36
(5)	Alpha Tm30					1	0.25	0.92	0.02	0.37	0.38	-0.07	-0.05	0.65	0.96	0.96	0.68	0.96	0.95
(6)	Beta Tm7						1	0.30	0.08	0.31	0.05	-0.15	-0.12	0.41	0.18	0.18	0.40	0.18	0.18
(7)	Beta Tm30							1	0.06	0.29	0.34	-0.13	-0.12	0.52	0.93	0.93	0.55	0.93	0.92
(8)	Downside Beta Tm30								1	-0.03	0.08	-0.12	-0.14	0.04	0.04	0.04	0.03	0.04	0.04
(9)	Coskew Tm30									1	0.02	0.06	0.10	0.73	0.19	0.19	0.73	0.19	0.19
(10)	iSkew Tm30										1	-0.15	-0.14	0.20	0.40	0.39	0.20	0.39	0.37
(11)	Shortfall 5% Tm7											1	0.94	-0.16	-0.13	-0.14	-0.12	-0.12	-0.11
(12)	VaR 5% Tm7												1	-0.11	-0.11	-0.12	-0.07	-0.09	-0.09
(13)	Vol Tm7													1	0.46	0.46	1.00	0.46	0.45
(14)	Vol Tm30														1	1.00	0.49	1.00	0.99
(15)	Vol Tm90															1	0.49	1.00	0.99
(16)	Ivol Tm7																1	0.49	0.48
(17)	Ivol Tm30																	1	0.99
(18)	Ivol Tm90																		1
	PC Financial	-0.03	-0.05	0.01	0.62	0.98	0.34	0.92	0.06	0.45	0.41	-0.19	-0.16	0.74	0.93	0.93	0.76	0.93	0.92

This table reports the correlation matrix among Financial Characteristics. Panel A reports pairwise Pearson correlation coefficients among the characteristics and the first principal component from them. The characteristics are re-scaled to be mean zero and unit variance before PCA and studying these correlations.

Table 2.19: Financial Characteristics: Signal.

Panel B. Financial characteristics exposures.

	0	7	14	30	90
Price	1.6e-07	-3.2e-08	-6.3e-07	-8.5e-07	-1.6e-06*
	(4.1e-07)	(2.6e-07)	(5.8e-07)	(6.8e-07)	(8.6e-07)
	0	0	0	0	0
Size	8.8e-03***	6.6e-04	4.4e-03***	5.6e-03*	4.7e-03
	(1.8e-03)	(1.2e-03)	(1.9e-03)	(2.9e-03)	(4.1e-03)
	0	0	0	0	0
MVRV	3.3e-03*	1.3e-03***	-5.5e-04	4.2e-04	-2.2e-03
	(1.9e-03)	(5.5e-04)	(1.5e-03)	(1.8e-03)	(1.8e-03)
	0	0	0	0	0
Alpha Tm7	208.74***	0.09	-1.3e+00	207.65***	208.82***
	(76.77)	(0.81)	(2.13)	(78.37)	(78.69)
	0.61	0	0	0.28	0.25
Alpha Tm30	142.39	3.1e-03	-3.7e+00	139.65	139.56
	(90.65)	(0.73)	(2.45)	(91.57)	(91.69)
	0.41	0	0	0.19	0.16
Beta Tm7	16.06	-7.2e-01***	-1.3e+00	16.03	16.94
	(13.47)	(0.20)	(1.61)	(13.92)	(14.27)
	0.15	0.01	0	0.07	0.07
Beta Tm30	26.33	-2.6e-01	-4.7e+00	23.89	24.11
	(21.42)	(0.20)	(3.33)	(21.93)	(21.98)
	0.25	0	0	0.1	0.09
Downside Beta Tm30	0.40	0.17***	0.99*	0.82	1.40***
	(0.33)	(0.05)	(0.51)	(0.50)	(0.52)
	0	0	0	0	0
Coskew Tm30	5.64**	-6.7e-02	-1.2e-01**	5.55**	5.62**
	(2.45)	(0.06)	(0.05)	(2.53)	(2.51)
	0.54	0	0	0.24	0.22
ISkew Tm30	0.12	1.8e-03	0.04	0.14	0.15
	(0.10)	(2.1e-03)	(0.03)	(0.11)	(0.12)
	0.03	0	0	0.02	0.02
Shortfall 5% Tm7	-2.1e-02	-7.5e-01***	-2.1e+01	-5.4e+00	-8.0e+00*
	(1.85)	(0.28)	(15.54)	(3.98)	(4.24)
	0	0	0.04	0	0.01
VaR 5% Tm7	3.91**	-1.6e+00***	-1.7e+01	-5.3e+00	-9.1e+00
	(1.99)	(0.39)	(10.72)	(6.85)	(7.18)
	0	0.01	0.01	0	0
Vol Tm7	19.69***	0.09***	0.84	19.82***	19.97***
	(0.36)	(0.03)	(1.20)	(0.26)	(0.16)
	0.94	0	0	0.45	0.4
Vol Tm30	6.63	-1.5e-02	0.34	6.81	6.87
	(5.94)	(0.05)	(0.34)	(5.96)	(5.97)
	0.19	0	0	0.09	0.08
Vol Tm90	9.91	0.02	0.45	10.24	10.40
	(8.86)	(0.08)	(0.41)	(8.89)	(8.90)
	0.19	0	0	0.09	0.09
Ivol Tm7	12.80***	0.05***	0.33	12.84***	12.90***
	(0.19)	(9.3e-03)	(0.51)	(0.17)	(0.13)
	0.95	0	0	0.45	0.4
Ivol Tm30	5.33	-1.4e-02	0.21	5.44	5.48
	(4.79)	(0.04)	(0.22)	(4.80)	(4.80)
	0.19	0	0	0.09	0.08
Ivol Tm90	9.81	-1.4e-02	0.41	10.05	10.08
	(8.80)	(0.07)	(0.41)	(8.83)	(8.85)
	0.19	0	0	0.09	0.08
PC Financial	0.27*	-2.0e-04	5.8e-03	0.27*	0.27*
	(0.14)	(1.3e-03)	(7.6e-03)	(0.14)	(0.14)
	0.5	0	0	0.24	0.21

This table reports the loadings on asset excess returns on each characteristic at various horizons. See the similar table descriptions for further detail.

Table 2.20: Principal Components of Characteristics: Correlations.

		(1)	(2)	(3)	(4)	(5)	(6)
(1)	Onchain	1	0.01	0.04	0.04	0.23	0.01
(2)	Exchange		1	0.36	0.02	0.53	-0.03
(3)	Social			1	0.06	0.44	0.01
(4)	Momentum				1	0.20	0.72
(5)	Microstructure					1	0.06
(6)	Financial						1

This table reports the correlation matrix among the first principal components of all groupings of asset characteristics, i.e. all pairwise Pearson correlation coefficients.

Table 2.21: Characteristic Signal by Year.

	2018	2019	2020	2021	2022	All
Onchain						
Tx Volume Tm7	0.025	0.021	0.015	0.007	0.007	0.005
Active Addresses Tm7	0.038	0.036	0.040	0.019	0.016	0.020
Δ Log New Addresses Tm14-Tm7	0.067	0.025	0.077	0.042	0.042	0.031
New Addresses Tm7	0.023	0.034	0.032	0.010	0.007	0.009
Total Addresses	0.047	0.036	0.031	0.014	0.013	0.021
Circulation Tm7	0.010	0.030	0.025	0.006	0.006	0.004
Age Destroyed	0.008	0.020	0.019	0.005	0.005	0.003
Δ Flow Distribution	0.115	0.063	0.029	0.029	0.057	0.044
Δ Holders Distribution	0.026	0.010	0.027	0.013	0.007	0.006
% Supply in Profit	0.104	0.054	0.089	0.075	0.082	0.080
Exchange						
% Circ. Supply CEX	0.028	0.026	0.059	0.037	0.038	0.034
% Circ. Supply DEX	0.050	0.011	0.049	0.019	0.019	0.018
% Circ. Supply Defi	0.050	0.068	0.017	0.002	0.002	0.001
% Circ. Supply Traders	0.031	0.030	0.056	0.044	0.048	0.041
Exchange Inflow Tm7	0.028	0.032	0.036	0.005	0.009	0.004
Exchange Outflow Tm7	0.020	0.028	0.036	0.004	0.008	0.004
Number of Trading Pairs	0.000	0.011	0.027	0.012	0.016	0.018
Social						
Social Volume	0.025	0.030	0.040	0.018	0.010	0.008
Social Volume Reddit	0.046	0.073	0.060	0.046	0.048	0.041
Social Volume Twitter	0.059	0.049	0.062	0.043	0.052	0.041
Sentiment Pos. Reddit	0.033	0.035	0.037	0.013	0.014	0.012
Sentiment Pos. Twitter	0.043	0.027	0.035	0.011	0.011	0.008
Sentiment Neg. Reddit	0.033	0.025	0.044	0.013	0.014	0.012
Sentiment Neg. Twitter	0.027	0.028	0.033	0.013	0.011	0.008
Developer Activity Tm7	0.042	0.051	0.078	0.040	0.039	0.037
VC Owned	0.008	0.006	0.010	0.003	0.003	0.002

This table reports, by year and overall, the pairwise mutual information between all weekly panel characteristics and asset excess returns seven days ahead.

Table 2.22: Characteristic Signal by Year (Continued).

Momentum						
Return Tm7	0.035	0.045	0.097	0.037	0.002	0.002
Return Tm14	0.079	0.046	0.085	0.040	0.003	0.004
Return Tm30	0.010	0.064	0.085	0.045	0.003	0.007
Return Tm60	0.030	0.050	0.082	0.038	0.003	0.013
Return Tm90	0.064	0.060	0.066	0.037	0.004	0.018
Return Tm14-Tm7	0.042	0.043	0.104	0.033	0.002	0.002
Return Tm30-Tm14	0.044	0.054	0.076	0.049	0.001	0.004
Return Tm90-Tm30	0.051	0.038	0.057	0.026	0.004	0.009
Return from ATH	0.072	0.043	0.100	0.067	0.054	0.056
Return from ATL	0.023	0.037	0.046	0.021	0.007	0.004
Return Industry Tm30	0.042	0.077	0.090	0.046	0.003	0.004
Return Industry Tm60	0.076	0.075	0.086	0.043	0.002	0.010
Microstructure						
Trades Sum Tm7	0.040	0.020	0.037	0.012	0.011	0.010
Volume Sum Tm7	0.028	0.025	0.026	0.016	0.015	0.012
Spread bps	0.034	0.031	0.018	0.019	0.001	0.011
Ask Size	0.000	0.126	0.151	0.193	0.048	0.101
Bid Size	0.000	0.074	0.072	0.036	0.014	0.018
Illiquidity Tm7	0.033	0.044	0.037	0.025	0.025	0.020
Turnover Tm7	0.023	0.025	0.033	0.007	0.008	0.005
Financial						
Price	0.035	0.033	0.026	0.011	0.014	0.012
Size	0.049	0.052	0.060	0.041	0.044	0.036
MVRV	0.059	0.043	0.059	0.006	0.008	0.007
Alpha Tm7	0.073	0.054	0.097	0.042	0.015	0.011
Alpha Tm30	0.097	0.078	0.085	0.050	0.003	0.003
Beta Tm7	0.070	0.061	0.068	0.031	0.027	0.017
Beta Tm30	0.072	0.059	0.044	0.033	0.011	0.011
Downside Beta Tm30	0.090	0.070	0.070	0.042	0.040	0.032
Coskew Tm30	0.054	0.059	0.061	0.033	0.012	0.008
ISkew Tm30	0.047	0.053	0.055	0.030	0.032	0.025
Shortfall 5% Tm7	0.047	0.049	0.060	0.071	0.034	0.035
VaR 5% Tm7	0.051	0.051	0.062	0.072	0.047	0.046
Vol Tm7	0.049	0.040	0.055	0.052	0.002	0.001
Vol Tm30	0.083	0.051	0.082	0.064	0.003	0.002
Vol Tm90	0.065	0.058	0.068	0.062	0.002	0.003
Ivol Tm7	0.052	0.050	0.053	0.053	0.002	0.002
Ivol Tm30	0.083	0.052	0.081	0.065	0.002	0.002
Ivol Tm90	0.063	0.056	0.082	0.058	0.002	0.002
T*N	410	449	671	4359	6894	

This table reports, by year and overall, the pairwise mutual information between all weekly panel characteristics and asset excess returns seven days ahead.

Table 2.23: Univariate Factor Returns: Statistically Significant Strategies.

	Quintiles					
	1	2	3	4	5	5-1
Return Tm14	-0.0031	-0.0012	0.0043	0.0093	0.0116	0.0147*
	-(0.29)	-(0.15)	(0.45)	(1.10)	(1.36)	(1.74)
Return Industry Tm30	0.0007	0.0025	0.0080	-0.0003	0.0122	0.0115*
	(0.09)	(0.27)	(0.90)	-(0.03)	(1.38)	(1.67)
Return Industry Tm60	0.0014	0.0026	0.0061	0.0027	0.0150	0.0136*
	(0.16)	(0.29)	(0.71)	(0.30)	(1.64)	(1.94)
Beta Tm7	0.0130	0.0099	0.0058	0.0062	-0.0026	-0.0156*
	(1.16)	(1.11)	(0.64)	(0.74)	-(0.29)	-(1.70)
iSkew Tm30	-0.0030	0.0033	0.0065	0.0032	0.0093	0.0123*
	-(0.40)	(0.40)	(0.72)	(0.30)	(1.02)	(1.77)
Shortfall5 Tm7	-0.0085	0.0074	0.0041	0.0061	0.0053	0.0138*
	-(0.78)	(0.72)	(0.48)	(0.68)	(0.80)	(1.69)

This table reports the mean quintile portfolio returns (and t-statistics) for characteristics with significant zero-investment strategies. The mean returns are the time-series averages of weekly value-weighted portfolio excess returns. 5-1 is the long-short top minus bottom quintile zero-investment portfolio. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 2.24: Univariate Factor Returns: Onchain Strategies.

	Quintiles					
	1	2	3	4	5	5-1
Tx Volume Tm7	0.0014	0.0048	0.0014	0.0082	0.0070	0.0056
	(0.15)	(0.52)	(0.16)	(0.89)	(0.87)	(0.94)
Active Addresses Tm7	0.0061	-0.0018	0.0075	0.0049	0.0060	-1e-04
	(0.60)	-(0.19)	(0.82)	(0.45)	(0.91)	-(0.01)
Δ Log New Addresses Tm14-Tm7	0.0025	0.0042	0.0109	0.0027	0.0050	0.0025
	(0.26)	(0.45)	(1.02)	(0.33)	(0.62)	(0.35)
New Addresses Tm7	0.0085	0.0017	0.0065	0.0047	0.0075	-0.0009
	(0.92)	(0.23)	(0.62)	(0.51)	(0.92)	-(0.15)
Total Addresses	0.0015	-0.0021	0.0119	0.0088	0.0052	0.0037
	(0.15)	-(0.23)	(1.16)	(0.77)	(0.78)	(0.52)
Circulation Tm7	0.0030	0.0131	0.0084	0.0008	0.0018	-0.0012
	(0.39)	(1.21)	(0.90)	(0.10)	(0.20)	-(0.18)
Age Destoryed	-0.0046	0.0076	0.0086	0.0099	0.0036	0.0082
	-(0.53)	(0.83)	(1.11)	(0.93)	(0.43)	(1.36)
Δ Flow Distribution	0.0046	0.0022	0.0076	-0.0023	0.0072	0.0026
	(0.55)	(0.25)	(0.83)	-(0.29)	(0.80)	(0.43)
Δ Holders Distribution	0.0119	0.0078	0.0001	0.0069	0.0014	-0.0104
	(1.15)	(0.91)	(0.01)	(0.77)	(0.17)	-(1.29)
% Supply in Profit	0.0016	0.0042	0.0049	0.0012	0.0064	0.0048
	(0.19)	(0.45)	(0.45)	(0.14)	(0.86)	(0.84)

This table reports the mean quintile portfolio returns (and t-statistics) for onchain characteristics. The mean returns are the time-series averages of weekly value-weighted portfolio excess returns. 5-1 is the long-short top minus bottom quintile zero-investment portfolio. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 2.25: Univariate Factor Returns: Exchange Strategies.

	Quintiles					
	1	2	3	4	5	5-1
% Circ. Supply CEX	0.0025	0.0073	0.0035	0.0065	0.0096	0.007
	(0.26)	(0.76)	(0.40)	(0.74)	(1.30)	(0.90)
% Circ. Supply DEX	0.0047	0.0025	0.0091	0.0057	0.0038	-0.0009
	(0.49)	(0.31)	(1.01)	(0.73)	(0.48)	-(0.10)
% Circ. Supply Defi	0.0025	0.0026	0.0045	0.0023	0.0094	0.0069
	(0.26)	(0.31)	(0.50)	(0.32)	(1.16)	(0.89)
% Circ. Supply Traders	0.0027	0.0003	0.0040	0.0097	0.0074	0.0047
	(0.29)	(0.03)	(0.46)	(1.10)	(1.00)	(0.59)
Exchange Inflow	0.0086	0.0097	0.0002	0.0146	0.0044	-0.0042
	(0.88)	(0.90)	(0.02)	(1.34)	(0.67)	-(0.64)
Exchange Outflow	0.0079	0.0096	-0.0003	0.0132	0.0045	-0.0034
	(0.81)	(0.90)	-(0.04)	(1.21)	(0.68)	-(0.52)
Number Trading Pairs	0.0069	0.0025	0.0057	0.0011	0.0057	-0.0012
	(0.71)	(0.28)	(0.61)	(0.13)	(0.73)	-(0.16)

This table reports the mean quintile sorted portfolio returns (and t-statistics) for exchange characteristics. The mean returns are the time-series averages of weekly value-weighted portfolio excess returns. 5-1 is the long-short top minus bottom quintile zero-investment portfolio. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 2.26: Univariate Factor Returns: Social Strategies.

	Quintiles					
	1	2	3	4	5	5-1
Social Volume	0.0073	0.0041	0.0032	0.0028	0.0057	-0.0015
	(0.73)	(0.42)	(0.30)	(0.30)	(0.87)	-(0.20)
Social Volume Reddit	-0.0011	0.0083	0.0007	0.0069	0.0057	0.0068
	-(0.12)	(0.85)	(0.08)	(0.67)	(0.86)	(1.07)
Social Volume Twitter	0.0013	0.0008	0.0072	0.0094	0.0034	0.0021
	(0.15)	(0.08)	(0.77)	(0.91)	(0.45)	(0.42)
Sentiment Pos. Reddit	0.0023	0.0036	0.0040	0.0062	0.0061	0.0037
	(0.23)	(0.40)	(0.43)	(0.56)	(0.91)	(0.50)
Sentiment Pos. Twitter	0.0038	0.0056	0.0057	0.0029	0.0060	0.0022
	(0.39)	(0.57)	(0.51)	(0.32)	(0.90)	(0.30)
Sentiment Neg. Reddit	0.0055	0.0048	0.0062	-0.0003	0.0056	1e-04
	(0.54)	(0.50)	(0.69)	-(0.04)	(0.85)	(0.02)
Sentiment Neg. Twitter	0.0096	0.0017	0.0075	0.0033	0.0064	-0.0032
	(0.96)	(0.16)	(0.69)	(0.37)	(0.95)	-(0.43)
Developer Activity	0.0085	-0.0029	0.0113	0.0031	0.0052	-0.0033
	(0.76)	-(0.33)	(1.12)	(0.35)	(0.78)	-(0.45)
VC Owned	0.0034				0.0052	0.0018
	(0.40)				(0.74)	(0.37)

This table reports the mean quintile portfolio returns (and t-statistics) for social characteristics. The mean returns are the time-series averages of weekly value-weighted portfolio excess returns. 5-1 is the long-short top minus bottom quintile zero-investment portfolio. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 2.27: Univariate Factor Returns: Momentum Strategies.

	Quintiles					
	1	2	3	4	5	5-1
Return Tm7	0.0002	0.0094	0.0084	0.0058	0.0079	0.0076
	(0.02)	(1.01)	(0.90)	(0.69)	(0.95)	(0.87)
Return Tm30	-0.0003	-0.0026	0.0004	0.0126	0.0060	0.0063
	-(0.03)	-(0.33)	(0.06)	(1.31)	(0.74)	(0.76)
Return Tm60	0.0057	0.0011	0.0101	0.0005	0.0092	0.0035
	(0.55)	(0.12)	(1.10)	(0.06)	(1.13)	(0.42)
Return Tm90	0.0040	0.0018	0.0048	0.0084	0.0033	-0.0006
	(0.36)	(0.20)	(0.53)	(0.96)	(0.42)	-(0.07)
Return Tm14-Tm7	-0.0002	-0.0020	0.0075	0.0075	0.0009	0.0011
	-(0.02)	-(0.25)	(0.86)	(0.90)	(0.11)	(0.14)
Return Tm30-Tm14	-0.0017	0.0010	0.0074	0.0157	-0.0010	0.0007
	-(0.16)	(0.12)	(0.86)	(1.66)	-(0.13)	(0.08)
Return Tm90-Tm30	0.0054	0.0050	0.0099	0.0070	-0.0043	-0.0098
	(0.53)	(0.55)	(0.94)	(0.95)	-(0.54)	-(1.28)
Return from ATH	0.0090	0.0024	0.0040	-0.0015	0.0027	-0.0063
	(0.72)	(0.28)	(0.47)	-(0.19)	(0.39)	-(0.65)
Return from ATL	0.0138	0.0049	0.0016	0.0059	0.0043	-0.0095
	(1.11)	(0.55)	(0.19)	(0.64)	(0.57)	-(1.00)

This table reports the mean quintile portfolio returns (and t-statistics) for momentum characteristics. The mean returns are the time-series averages of weekly value-weighted portfolio excess returns. 5-1 is the long-short top minus bottom quintile zero-investment portfolio. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 2.28: Univariate Factor Returns: Microstructure Strategies.

	Quintiles					
	1	2	3	4	5	5-1
Trades Sum Tm7	0.0005	0.0082	0.0048	0.0059	0.0047	0.0042
	(0.06)	(0.89)	(0.43)	(0.65)	(0.71)	(0.80)
Volume Sum Tm7	0.0068	0.0123	-0.0009	0.0083	0.0050	-0.0019
	(0.65)	(1.16)	-(0.10)	(0.90)	(0.74)	-(0.25)
Spread Bps	0.0060	-0.0030	0.0167	-0.0007	0.0049	-0.0011
	(0.91)	-(0.36)	(1.46)	-(0.07)	(0.49)	-(0.16)
Ask Size	0.0044	-0.0001	0.0046	0.0075	0.0081	0.0037
	(0.53)	-(0.02)	(0.50)	(0.82)	(0.99)	(0.67)
Bid Size	0.0054	0.0032	0.0008	0.0049	0.0070	0.0016
	(0.71)	(0.36)	(0.09)	(0.55)	(0.85)	(0.29)
Illiquidity Tm7	0.0059	0.0069	0.0052	0.0088	0.0015	-0.0044
	(0.89)	(0.80)	(0.56)	(0.82)	(0.15)	-(0.69)
Turnover Tm7	0.0005	0.0052	0.0052	0.0094	0.0048	0.0044
	(0.05)	(0.58)	(0.58)	(0.84)	(0.72)	(0.67)

This table reports the mean quintile portfolio returns (and t-statistics) for microstructure characteristics. The mean returns are the time-series averages of weekly value-weighted portfolio excess returns. 5-1 is the long-short top minus bottom quintile zero-investment portfolio. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 2.29: Univariate Factor Returns: Financial Strategies.

	Quintiles					
	1	2	3	4	5	5-1
Price	0.0063	0.0073	0.0037	0.0083	0.0059	-0.0004
	(0.63)	(0.76)	(0.41)	(0.80)	(0.87)	-(0.06)
Size	0.0060	0.0087	-0.0012	0.0085	0.0048	-0.0013
	(0.64)	(0.77)	-(0.13)	(0.88)	(0.72)	-(0.19)
MVRV	0.0086	0.0081	-0.0021	0.0059	0.0056	-0.0029
	(0.77)	(0.92)	-(0.22)	(0.74)	(0.77)	-(0.37)
Alpha Tm7	-0.0033	0.0008	0.0155	0.0002	0.0057	0.009
	-(0.31)	(0.10)	(1.46)	(0.03)	(0.65)	(1.07)
Alpha Tm30	-0.0038	-0.0032	0.0058	0.0071	0.0065	0.0103
	-(0.37)	-(0.40)	(0.73)	(0.80)	(0.75)	(1.25)
Beta Tm30	0.0091	0.0050	0.0096	0.0030	-0.0015	-0.0105
	(0.88)	(0.66)	(1.06)	(0.35)	-(0.15)	-(1.20)
Beta Downside Tm30	-0.0003	0.0101	0.0038	0.0043	0.0065	0.0068
	-(0.04)	(1.12)	(0.47)	(0.50)	(0.68)	(0.90)
Coskew Tm30	0.0025	0.0001	0.0064	0.0013	0.0098	0.0073
	(0.27)	(0.02)	(0.71)	(0.17)	(1.01)	(0.96)
VaR 5% Tm7	-0.0045	0.0085	-0.0008	0.0044	0.0049	0.0093
	-(0.40)	(0.86)	-(0.10)	(0.53)	(0.73)	(1.10)
Vol Tm7	0.0084	0.0034	0.0053	-0.0021	0.0027	-0.0058
	(1.23)	(0.38)	(0.61)	-(0.23)	(0.25)	-(0.73)
Vol Tm30	0.0072	0.0110	0.0032	-0.0019	0.0015	-0.0058
	(1.07)	(1.19)	(0.35)	-(0.21)	(0.14)	-(0.73)
Vol Tm90	0.0106	-0.0002	0.0087	-0.0012	0.0017	-0.0089
	(1.54)	-(0.03)	(0.90)	-(0.13)	(0.16)	-(1.04)
Ivol Tm7	0.0084	0.0043	0.0031	-0.0012	0.0035	-0.0049
	(1.23)	(0.47)	(0.36)	-(0.13)	(0.32)	-(0.62)
Ivol Tm30	0.0072	0.0104	0.0029	-0.0005	0.0010	-0.0062
	(1.07)	(1.13)	(0.32)	-(0.05)	(0.10)	-(0.79)
Ivol Tm90	0.0085	0.0028	-0.0024	0.0023	0.0051	-0.0034
	(1.07)	(0.37)	-(0.26)	(0.26)	(0.47)	-(0.36)

This table reports the mean quintile portfolio returns (and t-statistics) for financial characteristics. See the similar preceeding tables for further description.

Table 2.30: Low Dimensional Factor Model Out-of-Sample Returns: Multivariate, PCA, & IPCA.

Model	# Factors		Quintiles					5-1						
		Pred. R2	1	2	3	4	5	TS Avg	Sharpe	Sortino	Turnover	MDD	Alpha	Beta
Multi	1	< 0	-0.0127	-0.0121	-0.003	0.005	-0.0019	0.0107	1.31	2.74	0.06	-0.20	0.0103	-0.0441
			(-0.58)	(-0.66)	(-0.2)	(0.3)	(-0.09)	(0.93)					(0.0125)	(0.1544)
	2	< 0	-0.0101	-0.016	0.0029	-0.0073	0.0002	0.0103	1.72	3.87	0.06	-0.13	0.0106	-0.1145
			(-0.49)	(-0.87)	(0.21)	(-0.41)	(0.01)	(1.22)					(0.009)	(0.1106)
	3	< 0	-0.0129	-0.0128	-0.0027	0.0054	-0.002	0.011	1.44	2.94	0.06	-0.20	0.0099	-0.0614
			(-0.6)	(-0.7)	(-0.18)	(0.32)	(-0.1)	(1.02)					(0.0116)	(0.1424)
~ Fama	3	< 0	-0.0116	-0.0109	-0.0043	0.0032	-0.0063	0.0054	0.65	1.22	0.06	-0.20	0.0061	0.0662
			(-0.54)	(-0.62)	(-0.3)	(0.18)	(-0.28)	(0.46)					(0.0125)	(0.1544)
PCA	1	< 0	-0.0102	-0.0033	-0.0005	-0.0033	-0.0044	0.0058	0.78	1.41	0.06	-0.19	0.0053	0.0796
			(-0.47)	(-0.2)	(-0.03)	(-0.2)	(-0.2)	(0.55)					(0.0113)	(0.1394)
	2	< 0	-0.0054	-0.0147	-0.0007	-0.0076	-0.0045	0.0009	0.15	0.26	0.17	-0.13	0.0018	-0.0498
			(-0.26)	(-0.97)	(-0.05)	(-0.46)	(-0.23)	(0.11)					(0.0088)	(0.1088)
	3	< 0	-0.0108	-0.0123	-0.002	-0.0003	-0.0085	0.0023	0.45	1.01	0.21	-0.10	0.0037	0.0212
			(-0.52)	(-0.79)	(-0.13)	(-0.02)	(-0.4)	(0.32)					(0.0076)	(0.0939)
	4	< 0	-0.01	-0.0114	0.0013	-0.001	-0.0037	0.0063	1.34	3.18	0.25	-0.11	0.0059	0.0539
			(-0.54)	(-0.76)	(0.08)	(-0.06)	(-0.19)	(0.94)					(0.0072)	(0.0889)
	5	< 0	-0.017	-0.0111	-0.0031	0.0064	-0.0102	0.0068	1.24	2.45	0.29	-0.09	0.0065	0.0367
			(-0.85)	(-0.74)	(-0.21)	(0.35)	(-0.51)	(0.88)					(0.0084)	(0.1038)
IPCA	1	0.0002	-0.0188	-0.0087	-0.015	0.0003	0.0096	0.0284***	4.07	11.96	0.44	-0.08	0.0276***	-0.0234
			(-1.05)	(-0.53)	(-0.8)	(0.02)	(0.52)	(2.88)					(0.0107)	(0.1312)
	2	0.0014	-0.02	-0.0125	-0.0099	-0.0064	0.009	0.029**	3.46	11.22	0.42	-0.09	0.0274**	-0.1177
			(-1.02)	(-0.62)	(-0.62)	(-0.38)	(0.49)	(2.44)					(0.0126)	(0.1546)
	3	0.0018	-0.0206	-0.0145	-0.0086	0.0003	0.0103	0.0309**	3.59	11.13	0.41	-0.09	0.0301**	-0.3056**
			(-1.01)	(-0.73)	(-0.49)	(0.02)	(0.63)	(2.54)					(0.0121)	(0.1487)
	4	0.0004	-0.0227	-0.0139	-0.0113	0.0062	0.0013	0.024**	2.86	7.06	0.39	-0.09	0.0226*	-0.2846**
			(-1.09)	(-0.61)	(-0.76)	(0.33)	(0.09)	(2.02)					(0.0117)	(0.1446)
	5	-0.0014	-0.0139	-0.0172	-0.005	0.0002	0.0011	0.015	2.07	4.57	0.4	-0.13	0.0134	-0.1527
			(-0.72)	(-0.92)	(-0.3)	(0.01)	(0.06)	(1.47)					(0.0106)	(0.13)

This table reports—for multivariate factor models, PCA, and IPCA—the predictive R^2, the mean quintile portfolio returns, and portfolio statistics for the 5-1 strategy for July-December 2022, inclusive. For the 5-1 strategy, we report the time-series average weekly value-weighted excess return, annualized Sharpe Ratio, annualized Sortino, weekly turnover, maximum drawdown, and alpha and beta to the CMKT return. t-stats with *, **, and *** denote significance at the 10%, 5%, and 1% levels. For the multivariate factor model with 1, 2, and 3 factors, the selected characteristics are, respectively: size; illiquidity and size; and, size, 30 day momentum, and 90 day volatility.

Table 2.31: Univariate Factor Returns: Alpha and Loadings on Factor Model Strategies.

	alpha (1)	multi (2)	pca (3)	ipca (4)	R2 (5)
Return Tm14	-0.0089	-0.4051*	0.5248	0.2845	0.14
	(0.0108)	(0.2232)	(0.3613)	(0.1729)	
Return Industry Tm30	-0.0032	0.1016	-0.3278	-0.1324	0.08
	(0.0074)	(0.229)	(0.3102)	(0.0913)	
Return Industry Tm60	0.0117	0.0069	-0.1936	-0.1929	0.07
	(0.0078)	(0.2969)	(0.4608)	(0.1203)	
Beta Tm7	-0.0138	0.6266**	0.2789	-0.2806	0.18
	(0.0117)	(0.3159)	(0.6112)	(0.3301)	
iSkew Tm30	-0.0007	0.3981*	-0.0176	0.0359	0.11
	(0.0065)	(0.2256)	(0.2965)	(0.1883)	
Shortfall5 Tm7	-0.0138***	-0.2717	0.0284	-0.179	0.22
	(0.0049)	(0.1933)	(0.2301)	(0.2157)	

This table reports—for each univariate factor with a statistically significant 5-1 strategy—coefficients and standard errors from the contemporaneous time-series regression of the univariate factor 5-1 returns on the 5-1 returns for the best multivariate, PCA, and IPCA models. The best models were selected based on their Sharpe Ratio. The coefficients for alpha (i.e. intercept) and the three loadings are reported with standard errors in parentheses below. Standard errors are Newey-West adjusted using Bartlett's formula for the number of lags. *, **, and *** denote significance at the 10%, 5%, and 1% levels. The R^2 is reported in the last column.

Table 2.32: DSLFM Out-of-Sample Portfolio Statistics.

Weighting	# Factors		Quintiles					5-1						
		Pred. R2	1	2	3	4	5	TS Avg	Sharpe	Sortino	Turnover	MDD	Alpha	Beta
Mcap	1	0.0007	-0.0206	0.0025	-0.0008	-0.0069	-0.0051	0.0156	2.32	5.7	0.41	-0.14	0.0155	0.0309
			(-1.24)	(0.13)	(-0.05)	(-0.39)	(-0.3)	(1.64)					(0.0099)	(0.1239)
	2	< 0	-0.0106	-0.0	-0.0031	0.0044	-0.0149	-0.0043	-0.7	-0.96	0.28	-0.29	-0.005	0.3425***
			(-0.73)	(-0.0)	(-0.19)	(0.18)	(-0.76)	(-0.49)					(0.0072)	(0.0899)
	3	< 0	-0.0054	-0.0052	-0.0079	0.005	-0.0053	0.0001	0.01	0.03	0.37	-0.27	-0.0009	0.4762***
			(-0.39)	(-0.3)	(-0.44)	(0.19)	(-0.25)	(0.01)					(0.009)	(0.1133)
	4	< 0	-0.0107	-0.0057	0.0016	-0.0064	0.0033	0.014	1.83	4.02	0.33	-0.26	0.0133	0.3046**
			(-0.67)	(-0.38)	(0.08)	(-0.36)	(0.16)	(1.29)					(0.0101)	(0.1271)
	5	< 0	-0.0113	-0.0013	-0.0037	-0.0	-0.0024	0.0089	1.56	2.91	0.4	-0.16	0.0086	0.1257
			(-0.69)	(-0.08)	(-0.18)	(-0.0)	(-0.14)	(1.1)					(0.0081)	(0.1019)
Equal	1	0.0007	-0.0173	-0.0058	-0.016	-0.0122	-0.0046	0.0127**	3.36	8.18	0.41	-0.06	0.0125**	0.0825
			(-0.95)	(-0.32)	(-0.91)	(-0.65)	(-0.24)	(2.37)					(0.0054)	(0.068)
	2	< 0	-0.018	-0.0103	-0.0039	-0.0125	-0.0107	0.0073	1.44	2.19	0.28	-0.13	0.0065	0.3291***
			(-1.26)	(-0.54)	(-0.2)	(-0.64)	(-0.54)	(1.02)					(0.0051)	(0.0647)
	3	< 0	-0.0216	-0.0106	-0.0127	-0.0069	-0.0039	0.0177**	2.96	8.91	0.37	-0.06	0.0171**	0.3023***
			(-1.36)	(-0.6)	(-0.71)	(-0.35)	(-0.19)	(2.09)					(0.0073)	(0.0922)
	4	< 0	-0.0185	-0.0113	-0.0103	-0.0133	-0.0024	0.0161**	3.08	8.56	0.33	-0.08	0.0156**	0.2196**
			(-1.07)	(-0.66)	(-0.57)	(-0.73)	(-0.12)	(2.18)					(0.0068)	(0.0853)
	5	< 0	-0.016	-0.0028	-0.0125	-0.0156	-0.009	0.0069	1.05	2.71	0.4	-0.16	0.0066	0.1342
			(-0.9)	(-0.16)	(-0.7)	(-0.8)	(-0.45)	(0.74)					(0.0095)	(0.1191)

This table reports—for the DSLFM with market cap and equal-weighted portfolios—the predictive R^2, the mean quintile portfolio returns, and portfolio statistics for the 5-1 strategy for July-December 2022, inclusive. For each quintile, the mean returns are the time-series averages of weekly value-weighted portfolio excess returns sorted on each model's predicted returns. 5-1 is the long-short top minus bottom quintile zero-investment portfolio for each model; for which, we report the time-series average weekly value-weighted excess return, annualized Sharpe Ratio, annualized Sortino, weekly turnover, maximum drawdown, and alpha and beta to the CMKT return. t-stats are reported below each strategy's point estimates where *, **, and *** denote significance at the 10%, 5%, and 1% levels. Standard errors are Newey-West adjusted using Bartlett's formula for the number of lags.

Table 2.33: DSLFM: Asset Characteristic Significance.

	Estimate	Standard Error
Exchange Inflow	0.0558***	0.0161
Exchange Outflow	0.0547***	0.0161
Return Industry Tm30	0.0048	0.0046
Sentiment Neg. Reddit	0.0045	0.0071
Volume Sum Tm7	0.0044	0.0049
Alpha Tm7	0.0038	0.0043
Sentiment Pos. Reddit	0.0037	0.0058
Return Tm90	0.0026	0.0037
Social Volume Reddit	0.0022	0.0026
Alpha Tm30	0.0021	0.0021
Ask Size	0.0020	0.0026
Return Industry Tm60	0.0019	0.0023
Shortfall5 Tm7	0.0019	0.0021
Vol Tm90	0.0017	0.0018
Bid Size	0.0013	0.0018
% Supply in Profit	0.0011	0.0011
Active Addresses Tm7	0.0011	0.0014
Vol Tm7	0.0011	0.0020
Return Industry Tm30	0.0011	0.0016
Spread Bps	0.0009	0.0014
Return Tm7	0.0009	0.0011
Size	0.0009	0.0010
Illiquidity Tm7	0.0008	0.0014
Return Tm90-Tm30	0.0007	0.0010
Return from ATH	0.0007	0.0007
Return from ATL	0.0007	0.0008
Number of Trading Pairs	0.0006	0.0008
Vol Tm30	0.0006	0.0009
Turnover Tm7	0.0006	0.0010
Sentiment Neg. Twitter	0.0006	0.0010
Beta Tm7	0.0006	0.0009
Social Volume Twitter	0.0006	0.0010
Age Destoryed	0.0006	0.0008
Tx Volume Tm7	0.0006	0.0007

This table reports estimates of the importance of each asset characteristic to the fitted DSLFM using the test statistic $W_{\Gamma,j} = \Gamma_{\beta,j}^{\top}\Gamma_{\beta,j}$. The DSLFM latent loading estimate, i.e. $\widehat{\Gamma}_{\beta}$, comes from fitting the DSLFM to the entire weekly panel with hyperparameters selected by the DSLFM CV procedure for the best k, i.e. highest Sharpe. Standard errors are formed from the simulated distribution of $\widehat{W}_{\Gamma,j}$ using 200 bootstrap draws, for each j. *, **, and *** denotes significance at the 10%, 5%, and 1% levels. Only characteristics within two orders of magnitude of the maximum estimate are shown, i.e. 34 of the 63 characteristics.

CHAPTER 3

Doubly-Robust Inference for Conditional Average Treatment Effects with High-Dimensional Controls

This is joint work with Manu Navjeevan. See full paper for the full appendix with supporting mathematical proofs and empirical example.

3.1 Introduction

Consider a potential outcomes framework (Rubin 1974, 1978) where an observed outcome $Y \in$ and treatment $D \in \{0,1\}$ are related to two latent potential outcomes $Y_1, Y_0 \in$ via $Y = DY_1 + (1-D)Y_0$. To account for unobserved confounding factors a common strategy is to assume the researcher has access to a vector of covariates, $Z = (Z_1', X')' \in_1 \times \subseteq^{d_z - d_x} \times^{d_x}$, such that the potential outcomes are independent of the treatment decision after conditioning on the observed covariates, $(Y_1, Y_0) \perp D|Z$. In this setting, we are interested in estimation of and inference on the conditional average treatment effect (CATE):

$$\mathbb{E}[Y_1 - Y_0 \mid X = x]. \tag{3.1}$$

Estimation of the CATE generally requires first fitting propensity score and/or outcome regression models. When the number of control variables Z is large ($d_z \gg n$), these first-stage models must be estimated using regularized methods which converge slower than the nonparametric rate and typically rely on the correctness of parametric specifications for

consistency.[1]

Fortunately, if both models are correctly specified, one can obtain a nonparametric-rate consistent estimator and valid inference procedure for the CATE by using the popular augmented inverse propensity weighted (aIPW) signal (Semenova and Chernozhukov 2021b; Fan et al. 2022). This is because the aIPW signal obeys an orthogonality condition at, crucially, the true nuisance model values that limits the first-stage estimation error passed on to the second-stage estimator. Moreover, estimators based on the aIPW signal are doubly-robust; consistency of the resulting second-stage estimators requires correct specification of only one of the first-stage propensity score or outcome regression models. However inference based on these estimators is not doubly-robust. The orthogonality of the aIPW signal fails under misspecification and the resulting testing procedures and confidence intervals are rendered invalid.

This paper proposes a doubly-robust estimator and inference procedure for the conditional average treatment effect when the number of control variables d_z is potentially much larger than the sample size n. The dimensionality of the conditioning variable, d_x, remains fixed in our analysis. Our approach is based on Tan (2020) wherein doubly-robust inference is developed for the average treatment effect. We take a series approach to estimating the CATE, using a quasi-projection of the aIPW signal onto a growing set of basis functions. By assuming a logistic form for the propensity score model and a linear form for the outcome regression model, we construct novel ℓ_1-regularized first-stage estimating equations to recover a partial orthogonality of the aIPW signal at the limiting values of the first-stage estimators. So long as the limiting values of the first stage estimators have sparse representations this restricted orthogonality is enough to achieve doubly-robust pointwise and uniform inference; pointwise and uniform confidence intervals centered at the second-stage estimator are valid even if one of the logistic or linear functional forms is misspecified.

[1]Recent works by Bauer and Kohler (2019); Schmidt-Hieber (2020) provide some limited nonparametric results in high-dimensional settings using deep neural networks.

To achieve this restricted orthogonality at all points in the support of the conditioning variable, we employ distinct first-stage estimating equations for each basis term used in the second-stage series approximation. This results in the number of first-stage estimators growing with the number of basis terms. These estimators converge uniformly to limiting values under standard conditions in high-dimensional analysis. Improving on prior work in doubly-robust inference, our ℓ_1 regularized first-stage estimation incorporates a data-dependent penalty parameter based on the work of Chetverikov and Sω*rensen* (2021). This allows practical implementation of our proposed estimation procedure with minimal knowledge of the underlying data generating process.

The use of multiple pairs of nuisance parameter estimates limits our ability to straightforwardly apply existing nonparametric results for series estimators (Newey 1997; Belloni et al. 2015). Under modified conditions, we analyze the asymptotic properties of our second-stage series estimator to re-derive pointwise and uniform inference results. These modified conditions are in general slightly stronger than those of Belloni et al. (2015), though in certain special cases collapse exactly to the conditions of Belloni et al. (2015).

Prior Literature. Chernozhukov et al. (2018b) analyze the general problem of estimating finite dimensional target parameters in the presence of potentially high-dimensional nuisance functions. Using score functions that are Neyman-orthogonal with respect to nuisance parameters they show that it is possible to obtain target parameter estimates that are $\sqrt{n}$-consistent and asymptotically normal so long as the nuisance parameters are consistent at rate $n^{-1/4}$, a condition satisfied by many machine learning-based estimators. Semenova and Chernozhukov (2021b) take advantage of new results for series estimation in Belloni et al. (2015) and consider series estimation of functional target parameters after high-dimensional nuisance estimation.[2] The inference results of these papers are highly dependent on the orthogonality of their second stage estimators to first stage estimation error, making it difficult

[2]Fan et al. (2022) provides a similar analysis using a second-stage kernel estimator.

to directly extend these analyses when the first stage estimators are not consistent and the orthogonality cannot be applied.

In the same setting as this paper, Tan (2020) considers estimation of the average treatment effect. After assuming a logistic form for the propensity score and a linear form for the outcome regression, Tan (2020) proposes ℓ_1-regularized first-stage estimators that allow for partial control of the derivative of the aIPW signal away from true nuisance values and thus allow for doubly-robust inference. Smucler, Rotnitzky, and Robins (2019) extends the analysis of Tan (2020) to consider doubly-robust inference for a larger class of finite dimensional target parameters with bilinear influence functions. Wu et al. (2021) provide doubly-robust inference procedures for covariate-specific treatment effects with discrete conditioning variables; their results depend on exact representation assumptions that are unlikely to hold with continuous covariates. Moreover, no uniform inference procedures are described.

These papers pioneered the approach that we will employ below, which is to directly use the first order conditions of the first stage estimators to control second stage estimation error. However, it is not a priori clear how to extend this approach to control the estimation error passed onto an infinite dimensional target parameter like the CATE. As discussed above, our analysis requires re-deriving pointwise and uniform inference results for nonparametric series estimators under modified conditions.

Chetverikov and Sω*rensen* (2021) propose a data-driven "bootstrap after cross-validation" approach to penalty parameter selection that is modified for and implemented in our setting. This work is related to other work on the lasso (Tibshirani 1996; Bickel, Ritov, and Tsybakov 2009; Belloni and Chernozhukov 2013; Chetverikov, Liao, and Chernozhukov 2021) and ℓ_1-regularized M-estimation in high-dimensional settings (van der Greer 2016; Tan 2017).

Notation. For any measure F and any function f, define the L^2 norm, $\|f\|_{F,2} = (\mathbb{E}_F[f^2])^{1/2}$ and the L^∞ norm $\|f\|_{F,\infty} = \operatorname{ess\,sup}_F |f|$. For any vector in ^{p} let $\|\cdot\|_p$ for $p \in [1,\infty]$ denote the ℓ_p norm, $\|a\|_p = (\sum_{l=1}^p a_l^p)^{1/p}$ and $\|a\|_\infty = \max_{1\leq l\leq \infty} |a_l|$. If the subscript is unspecified,

we are using the ℓ_2 norm. For two vectors $a, b \in^p$, let $a \circ b = (a_i b_i)_{i=1}^p$ denote the Hadamard (element-wise) product. We adopt the convention that for $a \in^p$ and $c \in$, $a + c = (a_i + c)_{i=1}^p$. For a matrix $A \in^{m \times n}$ let $\|A\| = \max_{\|v\|_{\ell_2} \leq 1} \|Av\|_{\ell_2}$ denote the operator norm and $\|A\|_\infty = \sup_{1 \leq r \leq m, 1 \leq s \leq n} |A_{rs}|$. For any real valued function f let $\mathbb{E}_n[f(X)] = \frac{1}{n} \sum_{i=1}^n f(X_i)$ denote the empirical expectation and $\mathbb{G}_n[f(X)] = \frac{1}{\sqrt{n}} \sum_{i=1}^n (f(X_i) - \mathbb{E}[X_i])$ denote the empirical process. For two sequences of random variables $\{a_n\}$ and $\{b_n\}$, we say $a_n \lesssim_P b_n$ or $a_n = O_p(b_n)$ if a_n / b_n is bounded in probability and say $a_n = o_p(b_n)$ if $a_n / b_n \to_p 0$.

3.2 Setup

In this section, we formally define the setting and identification strategy that we consider. We then introduce our doubly-robust estimator and inference procedure. The parameter of interest is the conditional average treatment effect: $\mathbb{E}[Y_1 - Y_0 \mid X = x]$. However, for this paper we largely focus on estimation and inference for the conditional average counterfactual outcome:

$$g_0(x) \coloneqq \mathbb{E}[Y_1 \mid X = x]. \tag{3.2}$$

Doubly-robust estimation and inference on the other conditional counterfactual outcome, $\mathbb{E}[Y_0 \mid X = x]$, follows a similar procedure and is described in 3.5. The procedures can be combined for doubly-robust estimation and inference for the CATE.

3.2.1 Setting

We assume the researcher observes i.i.d data and conditioning on Z is sufficient to control for all confounding factors affecting both the treatment decision D and the potential outcomes, Y_1 and Y_0. Our analysis allows the dimensionality of the controls, $Z = (Z_1, X)$, to grow much faster than the sample size ($d_z \gg n$), while assuming the dimensionality of the conditioning variables, X, remains fixed ($d_x \ll n$).

Assumption 8 (Identification).

[(i)]$\{Y_i, D_i, Z_i\}_{i=1}^n$ *are independent and identically distributed.* $(Y_1, Y_0) \perp D \mid Z$*. There exists a value* $\eta \in (0,1)$ *such that* $\eta < \mathbb{E}[D \mid Z = z] < 1 - \eta$ *almost surely in* Z*.*

To obtain doubly-robust estimation and inference we use the augmented inverse propensity weighted (aIPW) signal,

$$Y(\pi, m) = \frac{DY}{\pi(Z)} - \left(\frac{D}{\pi(Z)} - 1\right) m(Z), \tag{3.3}$$

which is a function of a fitted propensity score model, $\pi(Z)$, and a fitted outcome regression model, $m(Z)$, whose true values are given $\pi^\star(Z) \coloneqq \mathbb{E}[D \mid Z]$ and $m^\star(Z) \coloneqq \mathbb{E}[Y \mid D = 1, Z]$. Under 8, the aIPW signal $Y(\cdot, \cdot)$ provides doubly-robust identification of $g_0(x)$. That is, for integrable $\pi \neq \pi^\star$ and $m \neq m^\star$,

$$\begin{aligned} \mathbb{E}[Y_1 \mid X = x] &= \mathbb{E}[Y(\pi^\star, m^\star) \mid X = x] \\ &= \mathbb{E}[\, Y(\pi, m^\star) \;\mid X = x] \\ &= \mathbb{E}[\, Y(\pi^\star, m) \;\mid X = x]. \end{aligned} \tag{3.4}$$

We use a series approach to estimate $g_0(x)$, taking a quasi-projection of the aIPW signal onto a growing set of k weakly positive basis terms:

$$p^k(x) \coloneqq (p_1(x), \ldots, p_k(x))' \in_+^k . \tag{3.5}$$

The basis terms are required to be weakly positive as they are used as weights within the convex first-stage estimators estimating equations.[3]Examples of weakly positive basis functions are B-splines or shifted polynomial series terms. To ensure that the basis terms are well behaved, we assume regularity conditions on $\xi_{k,\infty} \coloneqq \sup_{x\in} \|p^k(x)\|_\infty$, $\xi_{k,2} \coloneqq \sup_{x\in} \|p^k(x)\|_2$, and the eigenvalues of the design matrix $Q \coloneqq \mathbb{E}[p^k(x)p^k(x)']$.

For each basis term $p_j(x), j = 1, \ldots, k$, we estimate a separate propensity score model, $\widehat{\pi}_j(Z)$, and outcome regression model, $\widehat{m}_j(Z)$. Under standard moment and sparsity conditions, these converge uniformly over $j = 1, \ldots, k$ to limiting values $\bar{\pi}_j(Z)$ and $\bar{m}_j(Z)$. If the propensity score model and outcome regression models are correctly specified these limiting

values coincide with the true values $\pi^\star(Z)$ and $m^\star(Z)$. However, in general the limiting and true values may differ. The double robustness of the aIPW signal allows for identification of the CATE even if only one of the nuisance models is correctly specified. If either $\bar{\pi}_j = \pi^\star$ or $\bar{m}_j = m^\star$, we can write for all $j = 1, \ldots, k$:

$$\begin{aligned} Y(\bar{\pi}_j, \bar{m}_j) &= g_0(x) + \epsilon_j, \qquad & \mathbb{E}[\epsilon_j \mid X] = 0 \\ &= g_k(x) + r_k(x) + \epsilon_j \end{aligned} \tag{3.6}$$

where $g_0(x)$ is the conditional counterfactual outcome (3.2), $g_k(x) \coloneqq p^k(x)'\beta^k$ is the projection of $g_0(x)$ onto the first k basis terms, and $r_k(x) \coloneqq g_0(x) - g_k(x)$ denotes the approximation error from this projection. Note the separate error terms for each $j = 1, \ldots, k$ in (3.6), which are collected together in the vector $\epsilon^k \coloneqq (\epsilon_1, \ldots, \epsilon_k)$. As long as one of the first-stage models is correctly specified, the least squares parameter β^k governing the projection in $g_k(x)$ can be identified by the projection of the aIPW signal onto the basis terms $p^k(x)$:

$$\begin{aligned} \beta^k &\coloneqq Q^{-1}\mathbb{E}[p^k(X)Y_1] \\ &= Q^{-1}\mathbb{E}[p^k(X)Y(\pi^\star, m^\star)] \\ &= Q^{-1}\mathbb{E}[p^k(X)Y(\bar{\pi}_j, \bar{m}_j)], \quad \forall j = 1, \ldots, k. \end{aligned} \tag{3.7}$$

3.2.2 Estimator and Inference Procedure

We assume a logistic regression form for the propensity score model and a linear form for the outcome regression model:

$$\begin{aligned} \pi(Z; \gamma) &= (1 + \exp(-\gamma' Z))^{-1}, \\ m(Z; \alpha) &= \alpha' Z. \end{aligned} \tag{3.8}$$

[3] In case the researcher wants to use a second-stage basis that cannot be transformed to be weakly positive, we have shown a slightly modified method of constructing our doubly-robust estimator and inference procedure that does not require the first-stage weights to directly be the second-stage basis terms. This is available on request.

For each $j = 1, \ldots, k$, the parameters of (3.8), $\gamma, \alpha \in^{d_z}$, are estimated, respectively, by

$$\widehat{\gamma}_j \coloneqq \arg\min_{\gamma} \mathbb{E}_n[p_j(X)\{De^{-\gamma' Z} + (1-D)\gamma' Z\}] + \lambda_{\gamma,j}\|\gamma\|_1, \tag{3.9}$$

$$\widehat{\alpha}_j \coloneqq \arg\min_{\alpha} \mathbb{E}_n[p_j(X)De^{-\widehat{\gamma}_j' Z}(Y - \alpha' Z)^2]/2 + \lambda_{\alpha,j}\|\alpha\|_1. \tag{3.10}$$

The penalty parameters $\lambda_{\gamma,j}$ and $\alpha_{\gamma,j}$ are chosen via a data dependent technique described below. These first-stage estimating equations are designed so that their first order conditions directly limit the bias passed on to the second-stage series estimator, as is described in 3.3. Under standard assumptions the parameter estimators $\widehat{\gamma}_j, \widehat{\alpha}_j$ will converge uniformly over $j = 1, \ldots, k$ to population minimizers

$$\bar{\gamma}_j \coloneqq \arg\min_{\gamma} \mathbb{E}[p_j(X)\{De^{-\gamma' Z} + (1-D)\gamma' Z\}], \tag{3.11}$$

$$\bar{\alpha}_j \coloneqq \arg\min_{\alpha} \mathbb{E}[p_j(Z)De^{-\bar{\gamma}_j' Z}(Y - \alpha' Z)^2]. \tag{3.12}$$

which we assume are sufficiently sparse. Our first-stage estimators are then $\widehat{\pi}_j(Z) \coloneqq \pi(Z; \widehat{\gamma}_j)$ and $\widehat{m}_j(Z) \coloneqq m(Z; \widehat{\alpha}_j)$ with limiting values $\bar{\pi}_j(Z) \coloneqq \pi(Z; \bar{\gamma}_j)$ and $\bar{m}_j(Z) \coloneqq m(Z; \bar{\alpha}_j)$, respectively.

Our second-stage estimator $\widehat{g}(x) \coloneqq p^k(x)'\widehat{\beta}^k$ – where $\widehat{\beta}^k$ is an estimate of the population projection parameter β^k – is obtained by combining all k pairs of first-stage estimators

$$\widehat{\beta}^k \coloneqq \widehat{Q}^{-1}\mathbb{E}_n \begin{bmatrix} p_1(X)Y(\widehat{\pi}_1, \widehat{m}_1) \\ \vdots \\ p_k(X)Y(\widehat{\pi}_k, \widehat{m}_k) \end{bmatrix}, \tag{3.13}$$

and $\widehat{Q} \coloneqq \mathbb{E}_n[p^k(X)p^k(X)']$. We estimate the variance of $\widehat{g}(x)$ using $\widehat{\sigma}(x) \coloneqq \|\widehat{\Omega}^{1/2}p^k(x)\|/\sqrt{n}$ for

$$\widehat{\Omega} \coloneqq \widehat{Q}^{-1}\mathbb{E}_n[\{p^k(X) \circ \widehat{\epsilon}^k\}\{p^k(X) \circ \widehat{\epsilon}^k\}']\widehat{Q}^{-1}, \tag{3.14}$$

where $\circ$ represents the Hadamard product and $\widehat{\epsilon}^k \coloneqq (\widehat{\epsilon}_1, \ldots, \widehat{\epsilon}_k)$; $\widehat{\epsilon}_j \coloneqq Y(\widehat{\pi}_j, \widehat{m}_j) - \widehat{g}(x)$, $j = 1, ..., k$.

Inference is based on the $100(1-\eta)\%$ confidence bands

$$[\underline{i}(x), \bar{i}(x)] := [\widehat{g}(x) - c^\star(1-\eta/2)\,\widehat{\sigma}(x),\ \widehat{g}(x) + c^\star(1-\eta/2)\,\widehat{\sigma}(x)]\,. \tag{3.15}$$

For pointwise inference, the critical value $c^\star(1-\eta/2)$ is taken as the $(1-\eta/2)$ quantile of a standard normal distribution. For uniform inference $c^\star(1-\eta/2)$ is taken

$$c_u^\star(1-\eta/2) := (1-\eta/2)\text{-quantile of } \sup_{x\in} \left| \frac{p^k(x)\widehat{\Omega}^{1/2}}{\widehat{\sigma}(x)} N_k^b \right|$$

where N_k^b is a bootstrap draw from $N(0, I_k)$. Sections 3.3 and 3.4 show that, under standard sparsity and moment conditions, these pointwise and uniform inference procedures remain valid even under misspecification of either first-stage model.

3.2.3 Penalty Parameter Selection

To select the penalty parameters $\lambda_{\gamma,j}$ and $\lambda_{\alpha,j}$ in (3.9)-(3.10) we propose a data driven two-step procedure based on the work of Chetverikov and Sω*rensen* (2021). For each $j = 0, 1 \ldots, k$, we start with pilot penalty parameters given by

$$\lambda_{\gamma,j}^{\text{pilot}} = c_{\gamma,j} \times \sqrt{\frac{\ln^3(d_z)}{n}} \quad \text{and} \quad \lambda_{\alpha,j}^{\text{pilot}} = c_{\alpha,j} \times \sqrt{\frac{\ln^3(d_z)}{n}} \tag{3.16}$$

for some constants $c_{\gamma,j}, c_{\alpha,j}$ selected from the interval $[\underline{c}_n, \bar{c}_n]$ with $\underline{c}_n > 0$. In practice, the researcher has a fair bit of flexibility in choosing these constants. The optimal choice of these constants may depend on the underlying data generating process. We recommend using cross validation to pick these constants from a fixed-cardinality set of possible values. In line with 9(vi), the values in the set should be chosen to be on the order of the maximum value of $\|p^k(X_i)\|_\infty$ observed in the data.

Using $\lambda_{\gamma,j}^{\text{pilot}}$ and $\lambda_{\alpha,j}^{\text{pilot}}$ in lieu of $\lambda_{\gamma,j}$ and $\lambda_{\alpha,j}$ in (3.9)-(3.10) we generate pilot estimators $\widehat{\gamma}_j^{\text{pilot}}$ and $\widehat{\alpha}_j^{\text{pilot}}$. These pilot estimators are used to generate plug in estimators $\widehat{U}_{\gamma,j}$ and $\widehat{U}_{\alpha,j}$

of the residuals

$$\begin{aligned} \widehat{U}_{\gamma,j} &:= -p_j(X)\{De^{-\widehat{\gamma}_j^{\text{pilot}'}Z} + (1-D)\}U_{\gamma,j} \\ := -p_j(X)\{De^{-\gamma_j'Z} + (1-D)\} \\ \widehat{U}_{\alpha,j} &:= p_j(X)De^{-\widehat{\gamma}_j^{\text{pilot}'}Z}(Y - \widehat{\alpha}_j^{\text{pilot}'}Z). \end{aligned} \tag{3.17}$$

whose true values are given

$$\begin{aligned} U_{\gamma,j} &:= -p_j(X)\{De^{-\bar{\gamma}_j'Z} + (1-D)\} \\ U_{\alpha,j} &:= p_j(X)De^{-\bar{\gamma}'Z}(Y - \bar{\alpha}'Z) \end{aligned} \tag{3.18}$$

We then use a multiplier bootstrap procedure to select our final penalty parameters $\lambda_{\gamma,j}$ and $\lambda_{\alpha,j}$.

$$\begin{aligned} \lambda_{\gamma,j} &= c_0 \times (1-)\text{-quantile of } \max_{1\le l\le d_z} |\mathbb{E}_n[e_i\widehat{U}_{\gamma,j}Z_l]| \text{ given } \{Y_i, D_i, Z_i\}_{i=1}^n, \\ \lambda_{\alpha,j} &= c_0 \times (1-)\text{-quantile of } \max_{1\le l\le d_z} |\mathbb{E}_n[e_i\widehat{U}_{\alpha,j}Z_l]| \text{ given } \{Y_i, D_i, Z_i\}_{i=1}^n \end{aligned} \tag{3.19}$$

where $e_1, \ldots, e_n$ are independent standard normal random variables generated independently of the data $\{Y_i, D_i, X_i\}_{i=1}^n$ and $c_0 > 1$ is a fixed constant. In line with other work we find $c_0 = 1.1$ works well in simulations. So long as our residual estimates converge in empirical mean square to limiting values, the choice of penalty parameter in (3.19) will ensure that the penalty parameter dominates the noise with high probability. This allows for consistent variable selection and coefficient estimation.

3.3 Theory Overview

We begin with a main technical lemma which provides a bound on rate at which first-stage estimation error is passed on to the second-stage CATE and variance estimators. This bound is comparable to others seen in the inference after model-selection literature (Belloni, Chernozhukov, and Hansen 2013; Tan 2020) and is achieved under standard conditions in the ℓ_1-regularized estimation literature (Bickel, Ritov, and Tsybakov 2009; Bühlmann and van de

Geer 2011; Belloni and Chernozhukov 2013; Chetverikov and Sω*rensen* 2021). However, this bound is achieved at the limiting values of the propensity score and outcome regression models which may differ from the true values $\pi^\star$ and $m^\star$ under misspecification.

The potential misspecification of the first-stage models which means we cannot directly apply orthogonality of the aIPW signal, discussed below, to show that the effect of first-stage estimation error on the second-stage is negligible. Instead, we use the first order conditions for $\widehat{\gamma}_j$ and $\widehat{\alpha}_j$ to directly control this quantity. After presenting the lemma 3.3.2 provides some intuition for how this is done. Controlling the rate at which first-stage estimation error is passed on to the second-stage estimator even at points away from the true values $\pi^\star$ and $m^\star$ is key for obtaining doubly-robust inference for the CATE.

3.3.1 Uniform First-Stage Convergence

To show uniform convergence of the first-stage estimators and thus uniform control of the bias passed on from the first-stage estimation to the second-stage estimator we rely on the following assumption:

Assumption 9 (first-stage Convergence).

[(i)]The regressors Z are bounded, $\max_{1\leq l\leq d_z}|Z_l| \leq C_0$ almost surely. The errors $Y_1 - \bar{m}_j(Z)$ are uniformly subgaussian conditional on Z in the following sense. There exists fixed positive constants G_0 and G_1 such that for any j:

$$G_0\mathbb{E}\left[\exp\left(\{Y_1 - \bar{m}_j(Z)\}^2/G_0^2\right) - 1 \mid Z\right] \leq G_1^2$$

almost surely. There is a constant B_0 such that $\bar{\gamma}_j'Z \geq B_0$ almost surely for all j. There exists fixed constants $\xi_0 > 1$ and $1 > \nu_0 > 0$ such that for each $j = 1,\ldots,k$ the following empirical compatability condition holds for the empirical hessian matrix $\tilde{\Sigma}_{\gamma,j} := \mathbb{E}_n[De^{-\bar{\gamma}_j'Z}ZZ']$. For any $b \in^{d_z}$ and $_j = \{l : \bar{\gamma}_l \vee \bar{\alpha}_l \neq 0\}$:

$$\sum_{l\notin_j}|b_l| \leq \xi_0\sum_{l\in_j}|b_l| \implies \nu_0^2\Big(\sum_{l\in_j}|b_l|\Big)^2 \leq ||\left(b'\tilde{\Sigma}_{\gamma,j}b\right).$$

There exists fixed constants c_u *and* $C_U > 0$ *such that for all* $j = 1, \ldots, k$, $\mathbb{E}[U_{\gamma,j}^4] \leq (\xi_{k,\infty} C_U)^4$ *and* $\min_{1 \leq l \leq d_z} \mathbb{E}[U_{\gamma,j}^2 Z_l^2] \geq c_u$. *The constant* $\underline{c}_n$ *is chosen such that* $\xi_{k.\infty} \lesssim \underline{c}_n$ *and the following sparsity bounds hold for* $s_k = \max_{1 \leq j \leq k} |_j|$

$$\frac{\xi_{k,\infty} s_k^2 \bar{c}_n^2 \ln^5(d_z n)}{n} \to 0, \quad \textit{and} \quad \frac{\xi_{k,\infty}^4 \ln^7(d_z k n)}{n} \to 0.$$

The first part of 9 assumes that the regressors are bounded while the second assumes that tail behavior of the outcome regression errors are uniformly thin. Both of these can be relaxed somewhat with sufficient moment conditions on the tail behavior of the controls and errors. We should note that compactness of is generally required by nonparametric estimators. The third part of the assumption bounds all limiting propensity scores $\bar{\pi}_j(Z)$ away from zero uniformly. The fourth assumption is an empirical compatibility condition on the weighted first-stage design matrix. It is slightly weaker than the restricted eigenvalue conditions often assumed in the literature (Bickel, Ritov, and Tsybakov 2009; Belloni et al. 2012). The penultimate condition is an identifiability constraint that limits the moments of the noise and bounds it away from zero uniformly over all estimation procedures. Many of the constants in 9 are assumed to be fixed across all j. This is mainly to simplify the exposition of the results below and in practice all constants can be allowed to grow slowly with k. However, the growth rate of these terms affects the required first-stage sparsity.

The last condition is required for the validity of the bootstrap penalty parameter selection procedure and is comparable to the requirements needed for the bootstrap after cross validation technique described by Chetverikov and Sω*rensen* (2021). The main difference is the additional assumption on the growth rate of the basis functions, $\xi_{k,\infty}$ which is to ensure uniform stability of the estimation procedures (3.9)-(3.10) as well as some assumptions on the order of the constants $c_{\gamma,j}$ and $c_{\alpha,j}$ in (3.16).

LEMMA 14 (First-Stage Convergence). *Suppose that 9 holds. In addition assume that* $c_0 > (\xi_0 + 1)/(\xi_0 - 1)$, $k/n \to 0$, $k \to 0$, *and there is a fixed constant* $c > 0$ *such that for all* j, $\lambda_{\alpha,j}/\lambda_{\gamma,j} \geq c$.[4] *Then the following weighted means converge uniformly in absolute value at*

least at rate:

$$\max_{1\leq j\leq k} |\mathbb{E}_n[p_j(X)Y(\widehat{\pi}_j, \widehat{m}_j)] - \mathbb{E}_n[p_j(X)Y(\bar{\pi}_j, \bar{m}_j)]| \lesssim_P \frac{s_k\, \xi_{k,\infty}^2 \ln(d_z)}{n} \tag{3.20}$$

and in empirical mean square at least at rate:

$$\max_{1\leq j\leq k} \mathbb{E}_n[p_j^2(X)(Y(\widehat{\pi}_j, \widehat{m}_j) - Y(\bar{\pi}_j, \bar{m}_j))^2] \lesssim_P \frac{s_k^2\, \xi_{k,\infty}^4 \ln(d_z)}{n} \tag{3.21}$$

14 provides a tight bound on the first-stage estimation error passed on to the second-stage estimator even when the first-stage estimators converge to values that are not the true propensity score or outcome regression. In particular notice that under the (nearly familiar) sparsity bound $s_k \xi_{k,\infty}^2 k^{1/2} \ln^2(d_z)/\sqrt{n} \to 0$, any linear combination of the means in both (3.20) and (3.21) is $o_p(\sqrt{n})$. This allows us to obtain doubly-robust inference for the CATE.

3.3.2 Managing First-Stage Bias

We now provide some intuition for how this result is obtained and the role our particular estimating equations play in establishing this fact. We focus on control of the vector $\mathbf{B}^k$, defined in (3.22), which measures the bias passed on from first-stage estimation to the second-stage estimate $\widehat{\beta}^k$. Limiting the size of $\mathbf{B}^k$ is crucial in showing convergence of $\widehat{\beta}^k$ to the true parameter β^k and thus consistency of the nonparametric estimator $\widehat{g}(x)$.

$$\mathbf{B}^k := \mathbb{E}_n \begin{bmatrix} p_1(X)\left\{Y(\widehat{\pi}_1, \widehat{m}_1) - Y(\bar{\pi}_1, \bar{m}_1)\right\} \\ \vdots \\ p_k(X)\left\{Y(\widehat{\pi}_k, \widehat{m}_k) - Y(\bar{\pi}_k, \bar{m}_k)\right\} \end{bmatrix}. \tag{3.22}$$

For exposition, we consider a single term of (3.22), $\mathbf{B}_j^k$, which roughly measures the first-stage estimation bias taken on from adding the j^{th} basis term to our series approximation of

[4]The requirement $\lambda_{\alpha,j}/\lambda_{\gamma,j} \geq c$ may seem a bit unnatural, but it can be enforced in practice without upsetting any assumptions by setting the linear penalty $\lambda_{\alpha,j}^{\text{ratio}} := \max\{\lambda_{\gamma,j}/5, \lambda_{\alpha,j}\}$. In simulations, we find this constraint is rarely binding.

$g_0(x)$. The discussion that follows is a bit informal, instead of considering the derivatives with respect to the true parameters below our proof strategy will directly use the Kuhn-Tucker conditions of the optimization routines in (3.9)-(3.10). However, the general intuition is the same as is used in the proofs.

In addition to the doubly-robust identification property (3.4), the aIPW signal is typically useful in the high-dimensional setting because it obeys an orthogonality condition at the true values $(\pi^\star, m^\star)$:[5]

$$\mathbb{E}[\nabla_{\pi,m} Y(\pi^\star, m^\star) \mid Z] = 0. \tag{3.23}$$

When both the propensity score model and outcome regression model are correctly specified we can (loosely speaking) examine the bias $\mathbf{B}_j^k$ by replacing $\bar{\pi}_j = \pi^\star$ and $\bar{m}_j = m^*$ and considering the following first order expansion:

$$\begin{aligned}\mathbf{B}_j^k &= \mathbb{E}_n[p_j(X)Y(\widehat{\pi}_j, \widehat{m}_j)] - \mathbb{E}_n[p_j(X)Y(\pi^\star, m^\star)] \\ &= \underbrace{\mathbb{E}_n[p_j(X)\nabla_{\pi,m} Y(\pi^\star, m^\star)]}_{O_p(n^{-1/2}) \text{ by (3.23)}} \begin{bmatrix} \widehat{\pi}_j - \pi^\star \\ \widehat{m}_j - m^\star \end{bmatrix} + o_p(n^{-1/2}).\end{aligned} \tag{3.24}$$

By orthogonality of the aIPW signal the gradient term is close to zero, which guarantees that the bias is asymptotically negligible even if the nuisance parameters converge slowly to the true values, $\pi^\star$ and $m^\star$.[6] This allows the researcher to ignore first-stage nuisance parameter estimation error and treat $\pi^\star$ and $m^\star$ as known when analyzing the asymptotic properties of the second-stage series estimator. Indeed, since the aIPW signal orthogonality holds conditional on $Z = (Z_1, X)$, if both models are correctly specified only a single pair of first-stage estimators would be needed to provide control over all the elements in $\mathbf{B}^k$. This is the approach followed by Semenova and Chernozhukov (2021b).

[5] Robustness and orthogonality are indeed closely related, see Theorem 6.2 in Newey and McFadden (1994) for a discussion.

[6] Typically all that is required is that $\|\hat{\pi}_j - \pi^\star\| = o_p(n^{-1/4})$ and $\|\hat{m}_j - m^\star\| = o_p(n^{-1/4})$ in order to make the second order remainder term $\sqrt{n}$-negligible

So long as either one of $\bar{\pi}_j = \pi^\star$ or $\bar{m}_j = m^\star$, double robustness of the aIPW signal (3.4) still delivers identification: $\mathbb{E}[p_j(X)Y_1] \approx \mathbb{E}_n[p_j(X)Y(\bar{\pi}_j, \bar{m}_j)$. However, the aIPW orthogonality tells us nothing about the expectation of the gradient away from the true parameters, $\pi^\star, m^\star$; if either $\bar{\pi}_j \neq \pi^\star$ or $\bar{m}_j \neq m^\star$ there is no reason to believe that the gradient on the right hand side of (3.24) is mean zero when evaluated instead at $Y(\bar{\pi}_j, \bar{m}_j)$. In general, the bias $\mathbf{B}_j^k$ will then diminish at the rate of convergence of our nuisance parameters. Because we have high dimensional controls, this convergence rate will generally be much slower than the standard nonparametric rate (Newey 1997; Belloni et al. 2015).

To get around this, we design the first-stage objective functions (3.9)-(3.10) such that the resulting first-order conditions control the bias passed on to the second-stage. Consider the following expansion instead around the limiting parameters $\bar{\gamma}_j$ and $\bar{\alpha}_k$.

$$\begin{aligned}\mathbf{B}_j^k &= \mathbb{E}_n[p_j(X)Y(\widehat{\pi}_j, \widehat{m}_j)] - \mathbb{E}_n[p_j(X)Y(\bar{\pi}_j, \bar{m}_j)] \\ &= \mathbb{E}_n[p_j(X)\nabla_{\gamma_j,\alpha_j} Y(\bar{\pi}_j, \bar{m}_j)] \begin{bmatrix} \widehat{\gamma}_j - \bar{\gamma}_j \\ \widehat{\alpha}_j - \bar{\alpha}_j \end{bmatrix} + o_p(n^{-1/2})\end{aligned} \tag{3.25}$$

After substituting the forms of $\bar{\pi}_j(z) = \pi(z; \bar{\gamma}_j)$ and $\bar{m}_j(z) = m(z; \bar{\alpha}_j)$ described in (3.8) and differentiating with respect to γ_j and α_j we obtain

$$\mathbb{E}[p_j(X)\nabla_{\gamma_j,\alpha_j} Y(\bar{\pi}_j, \bar{m}_j)] = \mathbb{E}\begin{bmatrix} -p_j(X)De^{-\bar{\gamma}'Z}(Y - \bar{\alpha}'Z)Z \\ p_j(x)\{D(1 + e^{-\bar{\gamma}'Z})Z + Z\} \end{bmatrix} \tag{3.26}$$

However, by definition $\bar{\gamma}_j$ and $\bar{\alpha}_j$ solve the minimization problems defined in (3.11)-(3.12), the population analogs of our finite sample estimating equations. The first order conditions of these minimization problems yield

$$\mathbb{E}\underbrace{\begin{bmatrix} p_j(X)\{D(1 + e^{\bar{\gamma}'Z})Z + Z\} \\ p_j(X)De^{-\bar{\gamma}'Z}(DY - \bar{\alpha}'Z)Z \end{bmatrix}}_{\substack{\text{First order condition of } \bar{\gamma}_j \\ \text{First order condition of } \bar{\alpha}_j}} = 0 \implies \mathbb{E}[p_j(X)\nabla_{\gamma_j,\alpha_j} Y(\bar{\pi}_j, \bar{m}_j)] = 0 \tag{3.27}$$

Examining the first order conditions in (3.27), we see that they exactly give us control over the gradient (3.26). Under suitable convergence of the first-stage parameter estimates, this

guarantees the bias examined in expansion (3.25) is negligible even under misspecification of the propensity score or outcome regression models.

Control of this gradient under misspecification is not provided using other estimating equations, such as maximum likelihood for the logistic propensity score model or ordinary least squares for the linear outcome regression model. Moreover, control over the gradient of $\mathbf{B}_j^k$ from (3.22) is not provided by the first-order conditions for $\bar{\gamma}_l$ and $\bar{\alpha}_l$ for $l \neq j$:

$$\begin{aligned}\mathbb{E}[p_j(X)\nabla_{\gamma_j,\alpha_j} Y(\bar{\pi}_j, \bar{m}_j)] &= \mathbb{E}\overbrace{\begin{bmatrix} -p_j(X)De^{-\bar{\gamma}'Z}(Y - \bar{\alpha}'Z)Z \\ p_j(X)\{D(1+e^{\bar{\gamma}'Z})Z + Z\} \end{bmatrix}}^{\text{First order condition of } \bar{\gamma}_l} \\ &\neq \mathbb{E}\underbrace{\begin{bmatrix} p_l(X)\{D(1+e^{\bar{\gamma}'Z})Z + Z\} \\ p_l(X)De^{-\bar{\gamma}'Z}(Y - \bar{\alpha}'Z)Z \end{bmatrix}}_{\text{First order condition of } \bar{\alpha}_l}.\end{aligned} \tag{3.28}$$

Showing that the inference procedure of 3.2 remains valid at all points $x \in$ under misspecification requires showing negligible first-stage estimation bias for any linear combination of the vector (3.22). As outlined above, this requires using k separate pairs of nuisance parameter estimator to obtain k separate pairs of first order conditions, one for each term of the vector.

3.4 Main Results

In this section, we present the main consistency and distributional results for our second-stage estimator $\widehat{g}(x)$ described in 3.2. A full set of second-stage results, including pointwise and uniform linearization lemmas and uniform convergence rates, can be found in the Online Appendix. The first set of results is established under the following condition, which limits the bias passed from first-stage estimation onto the second-stage estimator. In particular, 3.4 implies that the bias vector $\mathbf{B}^k$ from (3.22) satisfies $\|\mathbf{B}^k\| = o_p(n^{-1/2})$.

[No Effect of First-Stage Bias]

$$\max_{1\leq j\leq k}\left|\mathbb{E}_n[p_j(X)Y(\widehat{\pi}_j,\widehat{m}_j)]-\mathbb{E}_n[p_j(X)Y(\bar{\pi}_j,\bar{m}_j)]\right|=o_p(n^{-1/2}k^{-1/2}). \tag{3.29}$$

Via 14 we can see that is a logistic propensity score model and a linear outcome regression model and estimating the first-stage models using the estimating equations (3.9)-(3.10), 3.4 can be achieved under 9 and the sparsity bound

$$\frac{s_k\,\xi_{k,\infty}^2 k^{1/2}\ln(d_z)}{\sqrt{n}}\to 0. \tag{3.30}$$

If the researcher were to assume different parametric forms for the first-stage model, different first estimating equations would have to be used to obtain doubly-robust estimation and inference. However, so long as the 3.4 can be established at the limiting values of the first-stage models, the results of this section hold.

Having dealt with the first-stage estimation error, the main complication remaining is that under misspecification the aIPW signals $Y(\hat{\pi}_j,\hat{m}_j)$ for $j=1,\ldots,k$ do not all converge to the same limiting values. However, so long as at least one of the first-stage models is correctly specified, all of the limiting aIPW signals have the same conditional mean, $g_0(x)$. In the standard setting, consistency of nonparametric estimator relies on certain conditions on the error terms. In our setting, we require that these assumptions hold uniformly over k the error terms. We note though that there is a non-trivial dependence structure between that limiting aIPW signals. This strong dependence gives plausibility to our uniform conditions. For example, if the logistic propensity score model is correctly specified and the limiting outcome regression models are uniformly bounded conditional on Z, our conditions reduce exactly to the conditions of Belloni et al. (2015). In general, however, the uniform conditions suggest that a degree of undersmoothing is optimal when implementing our estimation procedure.

3.4.1 Pointwise Inference

Pointwise inference relies on the following assumption in tandem with 3.4.

ASSUMPTION 10 (Second-Stage Pointwise Assumption). *Let* $\bar{\ }_k \coloneqq \max_{1\le j\le k} |_j|$. *Assume that*

[(i)]Uniformly over all n, *the eigenvalues of* $Q = \mathbb{E}[p^k(x)p^k(x)']$ *are bounded from above and away from zero. The conditional variance of the error terms is uniformly bounded in the following sense. There exists constants* $\underline{\sigma}^2$ *and* $\bar{\sigma}^2$ *such that for any* $j = 1, 2\ldots$ *we have that* $\underline{\sigma}^2 \le \mathrm{Var}(_j|\ X) \le \bar{\sigma}^2 < \infty$; *For each* n *and* k *there are finite constants* c_k *and* ℓ_k *such that for each* $f \in$

$$\|r_k\|_{L,2} = (\mathbb{E}[r_k(x)^2])^{1/2} \le c_k \quad \textit{and} \quad \|r_k\|_{L,\infty} = \sup_{x\in} |r_k(x)| \le \ell_k c_k.$$

$\sup_{x\in} \mathbb{E}[^2_k\, \mathbf{1}\{_k + \ell_k c_k > \delta\sqrt{n}/\xi_k\} \mid X = x] \to 0$ *as* $n \to \infty$ *and* $\sup_{x\in} \mathbb{E}[\ell_k^2 c_k^2 \mathbf{1}\{_k + \ell_k c_k > \delta\sqrt{n}/\xi_k\} \mid X = x] \to 0$ *as* $n \to \infty$ *for any* $\delta > 0$.

As mentioned, these are exactly the conditions required by Belloni et al. (2015), with the modification that the bounds on conditional variance and other moment conditions on the error term hold uniformly over $j = 1, \ldots, k$. The assumptions on the series terms being used in the approximation can be shown to be satisfied by a number of commonly used functional bases, such as polynomial bases or splines, under adequate normalizations and smoothness of the underlying regression function. Readers should refer to Newey (1997), Chen (2007), or Belloni et al. (2015) for a more in depth discussion of these assumptions.[7]

Under these assumptions, the variance of our second-stage estimator is governed by one of the following variance matrices:

$$\begin{aligned} \tilde{\Omega} &\coloneqq Q^{-1}\mathbb{E}[\{p^k(x) \circ (^k + r_k)\}\{p^k(x) \circ (^k + r_k)\}']Q^{-1} \\ \Omega_0 &\coloneqq Q^{-1}\mathbb{E}[\{p^k(x) \circ ^k\}\{p^k(x) \circ ^k\}']Q^{-1} \end{aligned} \tag{3.31}$$

where $\circ$ represents the Hadamard (element-wise) product and, abusing notation, for a vector $a \in^k$ and scalar $c \in$ we let $a + c = (a_i + c)_{i=1}^k$. Later on, we establish the validity of the plug-in analog $\hat{\Omega}$ (3.14), as an estimator of these matrices.

[7] In practice, we recommend the use of B-splines in order to to satisfy the first requirement that the basis functions are weakly positive and to reduce instability of the convex optimization programs described in (3.9)-(3.10).

THEOREM 2 (Pointwise Normality). *Suppose that 3.4 and 10 hold. In addition suppose that* $\xi_k^2 \log k/n \to 0$. *Then so long as either the logistic propensity score model or linear outcome regression model is correctly specified, for any* $\alpha \in S^{k-1}$:

$$\sqrt{n}\frac{\alpha'(\widehat{\beta}^k - \beta^k)}{\|\alpha'\Omega^{1/2}\|} \to_d N(0,1) \tag{3.32}$$

where generally $\Omega = \tilde{\Omega}$ *but if* $\ell_k c_k \to 0$ *then we can set* $\Omega = \Omega_0$. *Moreover, for any* $x \in$ *and* $s(x) \coloneqq \Omega^{1/2} p^k(x)$,

$$\sqrt{n}\frac{p^k(x)'(\widehat{\beta}^k - \beta^k)}{\|s(x)\|} \to_d N(0,1) \tag{3.33}$$

and if the approximation error is negligible relative to the estimation error, namely $\sqrt{n} r_k(x) = o(\|s(x)\|)$, *then*

$$\sqrt{n}\frac{\widehat{g}(x) - g(x)}{\|s(x)\|} \to_d N(0,1) \tag{3.34}$$

2 shows that the estimator proposed in 3.2 has a limiting gaussian distribution even under misspecification of either first-stage model. This allows for doubly-robust pointwise inference after establishing a consistent variance estimator.

3.4.2 Uniform Convergence

Next, we turn to strengthening the pointwise results to hold uniformly over all points $x \in$. This requires stronger conditions. we make the following assumptions on the tail behavior of the error terms which strengthens 10.

ASSUMPTION 11 (Uniform Limit Theory). *Let* $\bar{\ }_k = \sup_{1 \le j \le k} |_j|$, $\alpha(x) \coloneqq p^k(x)/\|p^k(x)\|$, *and let*

$$\xi_k^L \coloneqq \sup_{\substack{x,x' \in \\ x \neq x'}} \frac{\|\alpha(x) - \alpha(x')\|}{\|x - x'\|}.$$

Further for any integer s *let* $\bar{\sigma}_k^s = \sup_{x \in} \mathbb{E}[|\bar{}_k|^s | X = x]$. *For some* $m > 2$ *assume*

[(i)]The regression errors satisfy $\sup_{x \in} \mathbb{E}[\max_{1 \le i \le n} |\bar{}_{k,i}|^m \mid X = x] \lesssim_P n^{1/m}$ *The basis functions are such that (a)* $\xi_k^{2m/(m-2)} \log k/n \lesssim 1$, *(b)* $(\bar{\sigma}_k^2 \vee \bar{\sigma}_k^m) \log \xi_k^L \lesssim \log k$, *and (c)* $\log \bar{\sigma}_k^m \xi_k \lesssim \log k$.

As before, 11 is very similar to its analogue in Belloni et al. (2015), with the modification that the conditions are required to hold for $\bar{}_k$ as opposed to $_k$. Under this assumption, we derive doubly-robust uniform rates of convergence uniform inference procedures for the conditional counterfactual outcome $g_0(x)$.

THEOREM 3 (Strong Approximation by a Gaussian Process). *Assume that 3.4 holds and that Assumptions 10-11 hold with $m \geq 3$. In addition assume that (i) $\bar{R}_{1n} = o_p(a_n^{-1})$ and (ii) $a_n^6 k^4 \xi_k^2 (\bar{\sigma}_k^3 + \ell_k^3 c_k^2)^2 \log^2 n/n \to 0$ where*

$$\bar{R}_{1n} \coloneqq \sqrt{\frac{\xi_k^2 \log k}{n}}(n^{1/m}\sqrt{\log k} + \sqrt{k}\ell_k c_k) \quad \text{and} \quad \bar{R}_{2n} \coloneqq \sqrt{\log k} \cdot \ell_k c_k$$

Then so long as either the propensity score model or outcome regression model is correctly specified, for some $_k \sim N(0, I_k)$:

$$\sqrt{n}\frac{\alpha(x)'(\widehat{\beta} - \beta)}{\|\alpha(x)'\Omega^{1/2}\|} =_d \frac{\alpha(x)'\Omega^{1/2}}{\|\alpha(x)'\Omega^{1/2}\|}N_k + o_p(a_n^{-1}) \ \text{ in } \ell^\infty() \tag{3.35}$$

so that for $s(x) \coloneqq \Omega^{1/2} p^k(x)$

$$\sqrt{n}\frac{p^k(x)'(\widehat{\beta} - \beta)}{\|s(x)\|} =_d \frac{s(x)}{\|s(x)\|}N_k + o_p(a_n^{-1}) \ \text{ in } \ell^\infty() \tag{3.36}$$

and if $\sup_{x\in} \sqrt{n}|r_k(x)|/\|s(x)\| = o(a_n^{-1})$, then

$$\sqrt{n}\frac{\widehat{g}(x) - g(x)}{\|s(x)\|} =_d \frac{s(x)'}{\|s(x)\|_k} + o_p(a_n^{-1}) \ \text{ in } \ell^\infty() \tag{3.37}$$

where in general we take $\Omega = \tilde{\Omega}$ but if $\bar{R}_{2n} = o_p(a_n^{-1})$ then we can set $\Omega = \Omega_0$ where $\tilde{\Omega}$ and Ω_0 are as in (3.31).

3 establishes conditions under which we obtain a doubly-robust strong approximation of the empirical process $x \mapsto \sqrt{n}(\widehat{g}(x) - g_0(x))$ by a Gaussian process. After establishing consistent estimation of the matrix Ω, this strong approximation result allows us to show validity of the uniform confidence bands described in 3.2. As noted by Belloni et al. (2015), this is distinctly different from a Donsker type weak convergence result for the estimator $\widehat{g}(x)$ as viewed as a random element of $\ell^\infty(X)$. In particular, the covariance kernel is left completely unspecified and in general need not be well behaved.

3.4.3 Matrix Estimation and Uniform Inference

We establish that the estimator $\widehat{\Omega}$ proposed in (3.14) is a consistent estimator of the true limiting variance Ω, where $\Omega = \tilde{\Omega}$ in general but if $\bar{R}_{2n} = o_p(a_n^{-1})$ then $\Omega = \Omega_0$. To do so, we rely on the second-stage assumptions 10 and 11 as well as the following condition limiting the first-stage estimation error passed on to the variance estimator $\widehat{\Omega}$.

[Variance Estimation] Let $m > 2$ be as in 11. Then,

$$\xi_{k,\infty} \max_{1 \leq j \leq k} \mathbb{E}_n[p_j(X)^2 (Y(\widehat{\pi}_j, \widehat{m}_j) - Y(\bar{\pi}_j, \bar{m}_j))^2] = o_p(k^{-2} n^{-1/m}) \tag{3.38}$$

Via 14 we can establish 3.4.3 under 9 as well as the additional sparsity bound[8]

$$\frac{\xi_{k,\infty}^5 s_k^2 k^2 \ln(d_z)}{n^{(m-1)/m}}. \tag{3.39}$$

THEOREM 4 (Matrix Estimation). *Suppose that Conditions 3.4 and 3.4.3 and Assumptions 10-11 hold. In addition, assume that* $\bar{R}_{1n} + \bar{R}_{2n} \lesssim (\log k)^{1/2}$. *Then, so long as either the propensity score model or outcome regression model is correctly specified then for* $\widehat{\Omega} = \widehat{Q}^{-1} \widehat{\Sigma} \widehat{Q}^{-1}$:

$$\|\widehat{\Omega} - \Omega\| \lesssim_P (v_n \vee \ell_k c_k) \sqrt{\frac{\xi_k^2 \log k}{n}} = o(1)$$

4 establishes that pointwise inference based on the test statistic described in 3.2, obtained by replacing Ω in 2 with the consistent estimator $\widehat{\Omega}$, is doubly-robust. Hypothesis tests based on the test statistic as well as pointwise confidence intervals for $g_0(x)$ remain valid even if one of the first-stage parameters is misspecified.

We now establish the validity of uniform inference based on the gaussian bootstrap critical values $c_u^\star(1-\alpha)$ defined in 3.2.

[8] The sparsity bound (3.39) required for consistent variance estimation can be significantly sharpened if the researcher is willing to use a cross fitting procedure, using one sample to estimate the nuisance parameters and another to evaluate the aIPW signal. This is because one could more directly follow Semenova and Chernozhukov (2021b) and control alternate quantities with bounds that converge more quickly to zero.

THEOREM 5 (Validity of Uniform Confidence Bands). *Suppose* **??** *are satisfied and assm:second-stage-assumptionsassm:uniform-limit-theory hold with* $m \geq 4$. *In addition suppose (i)* $R_{1n} + R_{2n} \lesssim \log^{1/2} n$, *(ii)* $\xi_k \log^2 n / n^{1/2-1/m} = o(1)$, *(iii)* $\sup_{x\in} |r_k(x)|/\|p^k(x)\| = o(\log^{-1/2} n)$, *and (iv)* $k^4 \xi_k^2 (1 + l_k^3 r_k^3)^2 \log^5 n / n = o(1)$. *Then, so long as either the propensity score model or outcome regression model is satisfied*

$$\Pr\left(\sup_{x\in} |\frac{\widehat{g}(x) - g(x)}{\widehat{\sigma}(x)}| \leq c^\star(1-\alpha)\right) = 1 - \alpha + o(1).$$

As a result, uniform confidence intervals formed in (3.15) *satisfy*

$$\Pr(g(x) \in [\underline{i}(x), \bar{i}(x)],\ \forall x \in) = 1 - \alpha + o(1).$$

In conjunction with 14, 2 and 4, 5 shows the validity of the uniform inference procedure described in 3.2.

3.5 Estimation of the Conditional Average Treatment Effect

Up to now, we have mainly focused on doubly-robust estimation and model-assisted inference for the function

$$g_0(x) = \mathbb{E}[Y_1 \mid X = x].$$

We conclude by noting that we can use a symmetric procedure to obtain model-assisted inference for the additional conditional counterfactual outcome

$$\tilde{g}_0(x) = \mathbb{E}[Y_0 \mid X = x].$$

To do so, we use the alternate aIPW signal

$$Y_0(\pi_0, m_0) = \frac{(1-D)Y}{1-\pi_0(Z)} + \left(\frac{1-D}{1-\pi_0(Z)} - 1\right) m_0(Z)$$

where as before the true value for $\pi_0^\star(z) = \Pr(D = 1 \mid Z = z)$ but now $m_0^\star(z) = \mathbb{E}[Y \mid D = 0, Z = z]$. To estimate these nuisance models we again assume a logistic form for the

propensity score model $\pi_0(z) = \pi(z; \gamma^0)$ and a linear form for the outcome regression model $m_0(z) = m(z, \alpha^0)$ as in (3.8) and use a separate estimation procedure for each basis term in our series approximation of $\tilde{g}_0(x)$. The estimating equations we use to estimate each γ_j^0 and α_j^0 differ from those in (3.9)-(3.10) however, and are instead given

$$\widehat{\gamma}_j^0 := \arg\min_{\gamma} \mathbb{E}_n[p_j(X)\{(1-D)e^{\gamma' Z} - D\gamma' Z\}] + \lambda_{\gamma,j}\|\gamma\|_1$$
$$\widehat{\alpha}_j^0 := \arg\min_{\alpha} \mathbb{E}_n[p_j(Z)(1-D)e^{\widehat{\gamma}_j^{0\prime} Z}(Y - \alpha' Z)^2]/2 + \lambda_{\alpha,j}\|\alpha\|_1$$

which under the natural analog of 9 converge uniformly to population minimizers:

$$\bar{\gamma}_j^0 := \arg\min_{\gamma} \mathbb{E}[p_j(X)\{(1-D)e^{\gamma' Z} - D\gamma' Z\}]$$
$$\bar{\alpha}_j^0 := \arg\min_{\alpha} \mathbb{E}[p_j(Z)(1-D)e^{\bar{\gamma}_j^{0\prime} Z}(Y - \alpha' Z)^2]$$

Letting $\bar{\pi}_{0,j}(z) = \pi(z, \bar{\gamma}_j^0)$, and $\bar{m}_{0,j}(z) = m(z, \bar{\alpha}_j^0)$ we can repeat the decomposition of 3.3, expressing $\tilde{Y}(\bar{\pi}_{0,j}, \bar{m}_{0,j})$ as functions of the parameters $\bar{\gamma}_j^0$ and $\bar{\alpha}_j^0$ and show that the first order conditions for $\bar{\gamma}_j^0$ and $\bar{\alpha}_j^0$ directly control the bias passed on to the second stage nonparametric estimator for $\tilde{g}_0(x)$. Convergence rates and validity of inference then follow from symmetric analysis of the results in **??**. Combining estimation and inference of the two conditional counterfactual outcomes then gives a doubly-robust estimator and inference procedure for the CATE. To perform inference on the CATE we can use the variance matrix

$$\bar{\Omega} = \Omega_0 + \Omega_1 - 2\Omega_2$$

where Ω_0 is as in (3.31) but Ω_1 and Ω_2 are given

$$\begin{aligned} \Omega_1 &= Q^{-1}\mathbb{E}[\{p^k(x)\circ_0^k\}\{p^k(x)\circ_0^k\}']Q^{-1} \\ \Omega_2 &= Q^{-1}\mathbb{E}[\{p^k(x)\circ^k\}\{p^k(x)\circ_0^k\}']Q^{-1} \end{aligned} \tag{3.40}$$

where ${}_{0,j}^{k} = Y_0(\bar{\pi}_{0,j}, \bar{m}_{0,j}) - \tilde{g}_0(x)$ and ${}_{0}^{k} = ({}_{0,1}^{k}, \ldots, {}_{0,k}^{k})'$. These matrices can be consistently estimated using their natural empirical analogs as in (3.14).

3.6 Simulation Study

We investigate the finite-sample performance of the doubly-robust estimator and inference procedure via simulation study. We find that our proposed estimation procedure retains good coverage properties even under misspecification.

3.6.1 Simulation Design

Observations are generated i.i.d. according to the following distributions The error term is generated following $\epsilon \sim N(0,1)$. The controls are set $Z_i = (Z_{1i}, X_i) \in^{d_z}$ where $d_z = 100$, $X \sim U(1,2)$, and the independent regressors Z_1 are jointly centered Gaussian with a covariance matrix of the Toeplitz form

$$\text{Cov}(Z_{1,j}, Z_{1,k}) = \mathbb{E}[Z_{1,j} Z_{1,k}] = 2^{-|j-k|}, \quad 3 \leq j, k \leq d_z.$$

To capture misspecification, we let $Z^\dagger$ be a transformation of the regressors in Z_1 where $Z_j^\dagger = Z_j + \max(0, 1 + Z_j)^2, \ \forall \ j = 3, \ldots, d_z$. Let `sparsity` control the number of regressors in $Z = (Z_1, X)$ entering the DGP.

[label=(S0)] *Correct specification*: Generate D given Z from a Bernoulli distribution with $\Pr(D = 1|Z) = \{1 + \exp(p_1 - X - 0.5X^2 - \gamma' Z_1)\}^{-1}$ and $Y = D(1 + X + 0.5X^2 + \gamma' Z_1) + \epsilon$. *Propensity score model correctly specified, but outcome regression model misspecified*: Generate D given Z as in (S1), but $Y = D(1+X+0.5X^2+\gamma' Z_1^\dagger)+\epsilon$. *Propensity score model misspecified, but outcome regression model correctly specified*: Generate Y according to (S1), but generate D given Z from a Bernoulli distribution with $\Pr(D = 1|Z) = \{1 + \exp(p_2 - X - 0.5X^2 + \gamma' Z_1^\dagger)\}^{-1}$.

where the constants p_1 and p_2 differ in various simulation setups but are always set so that the average probability of treatment is about one half. To consider various degrees of high-dimensionality, we implement $N \in \{500, 1000\}$ with $d_z = 100$. For (S1), `sparsity`= 6; for (S2), `sparsity`= 4; and, for (S3), `sparsity`= 5. Results are reported for $S = 1,000$

repeated simulations.

3.6.2 Estimators and Implementation

To select the first stage penalty parameters, we implement the multiplier bootstrap procedure described in 3.2.3. The constants $c_{\gamma,j}$ and $c_{\alpha,j}$ in the pilot penalty parameters (3.16) are selected via cross validation from a set of size 5. To select the final bootstrap penalty parameter we set $c_0 = 1.1$ and select the 95^{th} quantile of $B = 10000$ bootstrap replications. In our second-stage estimation, we use a b-spline basis of size $k = 3$. B-splines are implemented from the R package `splines2` (Wang and Yan 2021), which uses the specification detailed in Perperoglou et al. (2019). In the tables below, we refer to our method as *MA-DML* (model assisted double machine learning).

We compare our proposed estimator and inference procedure to that of Semenova and Chernozhukov (2021b), which projects a single aIPW signal onto a growing series of basis terms. In implementing this *DML* method, we use the standard ℓ_1-penalized maximum likelihood (MLE) and ordinary least squares (OLS) loss functions to estimate the first stage propensity score and outcome regression models, respectively.

Estimation error is studied for the target parameter $g_0(x) = \mathbb{E}[Y|D = 1, X = x]$ over a grid of 100 points spaced across $x \in [1, 2]$, i.e. the support of X. We study average coverage across simulations of each method's pointwise (at $x = 1.5$) and uniform confidence intervals. To compare the estimation error for the target parameter $g(x)$ across the two different estimators $\widehat{g}_s(x)$ for each simulation $s = 1, \ldots, S$, we utilize integrated bias, variance, and

mean-squared error where $\bar{g}(x) = S^{-1}\sum_{s=1}^{S}\widehat{g}_s(x)$,

$$\begin{aligned}
\text{IBias}^2 &= \int_0^1 (\bar{g}(x) - g_0(x))^2 dx, \\
\text{IVar} &= S^{-1}\sum_{s=1}^{S}\int_0^1 (\widehat{g}_s(x) - \bar{g}(x))^2 dx, \\
\text{IMSE} &= S^{-1}\sum_{s=1}^{S}\int_0^1 (\widehat{g}_s(x) - g_0(x))^2 dx.
\end{aligned}$$

3.6.3 Simulation Results

Table 3.1 presents the simulation results for all three specifications (S1)-(S3) for $n = 500$ and $n = 1000$. Integrated squared bias, variance, and mean squared error are presented in columns (1)-(3), respectively. Pointwise and uniform coverage results are presented in columns (4)-(7).

For pointwise and uniform coverage under correct specification regime (S1), *MA-DML* has some slight improvements. Under misspecification DGPs (S2) and (S3), the pointwise coverage of *MA-DML* is closer to the targets except in the $N = 1000$ and (S2) case where it slightly underperforms. However, *MA-DML* has a notable improvement over *DML* in the (S3) case when $N = 1000$. Similarly, *MA-DML* outperforms *DML* in three of the four misspecified regimes, i.e. all but (S3) when $N = 500$ where *MA-DML* has over-coverage. Under (S2) when $N = 1000$, both methods are markedly deterioated uniform coverage, although *MA-DML* is noticably closer to target.

In regards to estimation error, in four of the six settings, *MA-DML* has a lower MSE than *DML* where regardless of sample size *MA-DML* underperforms in (S3). Notably, it does appear *MA-DML* has substantially smaller IBias^2 across the DGPs.

Finally, we were surprised to find for both estimators that coverage properties, in general, improve under the higher-dimensional regime of $N = 500$ with $d_z = 100$ compared to

Table 3.1: Simulation study.

DGP	Estimator	IBias^2 (1)	IVar (2)	IMSE (3)	Cov90 (4)	Cov95 (5)	UCov90 (6)	UCov95 (7)
		K=3, n=500, $d_z = 100$						
(S1)	DML	0.04	0.31	0.35	0.92	0.96	1.00	1.00
	MA-DML	$\sim$0.0	0.34	0.34	0.93	0.97	1.00	1.00
(S2)	DML	0.16	2.17	2.33	0.92	0.97	0.83	0.86
	MA-DML	0.03	2.12	2.15	0.90	0.94	0.88	0.91
(S3)	DML	0.03	0.55	0.59	0.87	0.93	0.95	0.97
	MA-DML	0.01	0.79	0.80	0.91	0.95	0.99	0.99
		K=3, n=1000, $d_z = 100$						
(S1)	DML	0.12	0.20	0.32	0.83	0.90	0.96	0.96
	MA-DML	0.01	0.22	0.23	0.83	0.90	0.99	0.99
(S2)	DML	0.40	2.1	2.5	0.84	0.91	0.33	0.39
	MA-DML	0.19	2.07	2.26	0.83	0.89	0.50	0.55
(S3)	DML	0.11	0.34	0.46	0.74	0.82	0.80	0.84
	MA-DML	0.01	0.53	0.54	0.84	0.89	0.89	0.91

Note: DGP refers to the three various data generating processes introduced above. IBias^2, IVar, and IMSE refer to integrated squared bias, variance, and mean squared error, respectively. Cov90, Cov95, UCov90, and UCov95 refer to the coverage proportion of the 90% and 95% pointwise and uniform confidence intervals across simulations. K refers to the number of series terms, N to the sample size, and d_z to the dimensionality of the random variable Z_1.

$N = 1,000$ and $d_z = 100$. In particular, with a higher ratio of covariates to observations, the uniform coverage properties under regime (S2) were substantially better. The estimation error results were in line with our priors as the higher-dimensional regime sees in general higher estimation errors for both methods.

For coverage under correct specification, we did anticipate the underperformance of *MA-DML* given it is designed to handle misspecification with the cost of other estimators outperforming under correct specification. Additionally, we attribute the poor uniform coverage in DGP (S2) for both estimators under $N = 1,000$ to a lack of a rich enough cross-validation given the performance was improved under a more difficult regime when the number of observations drops to $N = 500$. The integrated bias of *MA-DML* is lower across the various DGPs compared to *DML*. Following the discussion in 3.3 this is expected since the first stage estimating equations for the model assisted procedure are specifically designed to minimize the bias passed on to the second stage estimator. However, the model assisted procedure has higher values of integrated variance compared to the standard procedure, which could be attributable to the use of k distinct first-stage estimations.

Our findings should not be interpreted as a critique of the Semenova and Chernozhukov (2021b) benchmark method, whose work we rely on and were inspired by.

3.7 Conclusion

Estimation of conditional average treatment effects with high dimensional controls typically relies on first estimating two nuisance parameters: a propensity score model and an outcome regression model. In a high-dimensional setting, consistency of the nuisance parameter estimators typically relies on correctly specifying their functional forms. While the resulting second-stage estimator for the conditional average treatment effect typically remains consistent even if one of the nuisance parameters is inconsistent, the confidence intervals may no longer be valid.

In this paper, we consider estimation and valid inference on the conditional average treatment effect in the presence of high dimensional controls and nuisance parameter misspecification. We present a nonparametric estimator for the CATE that remains consistent at the nonparametric rate, under slightly modified conditions, even under misspecification of either the logistic propensity score model or linear outcome regression model. The resulting Wald-type confidence intervals based on this estimator also provide valid asymptotic coverage under nuisance parameter misspecification.

Bibliography

Aiello, Darren, Scott R Baker, Tetyana Balyuk, Marco Di Maggio, Mark J Johnson, and Jason D Kotter. 2023. "Who Invests in Crypto? Wealth, Financial Constraints, and Risk Attitudes."Technical report, National Bureau of Economic Research.

Ang, Andrew, Jun Liu, and Krista Schwarz. 2009. "Using individual stocks or portfolios in tests of factor models." *AFA 2009 San Francisco Meetings Paper*.

Babiak, Mykola, and Daniele Bianchi. 2021. "A risk-based explanation of cryptocurrency returns." *Available at SSRN*.

Bai, Jushan. 2003. "Inferential theory for factor models of large dimensions." *Econometrica* 71 (1): 135–171.

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

Bai, Jushan, and Serena Ng. 2013. "Principal components estimation and identification of static factors." *Journal of econometrics* 176 (1): 18–29.

Bali, Turan G, Robert F Engle, and Scott Murray. 2016. *Empirical asset pricing: The cross section of stock returns.*: John Wiley & Sons.

Bauer, Benedikt, and Michael Kohler. 2019. "On deep learning as a remedy for the curse of dimensionality in nonparametric regression." *The Annals of Statistics* 47 (4): 2261 – 2285.

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen. 2012. "Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain." *Econometrica* 80 (6): 2369–2429.

Belloni, Alexandre, and Victor Chernozhukov. 2013. "Least squares after model selection in

high-dimensional sparse models." *Bernoulli* 19 (2): 521 – 547.

Belloni, Alexandre, Victor Chernozhukov, Denis Chetverikov, Christian Hansen, and Kengo Kato. 2018. "High-dimensional econometrics and regularized GMM." *arXiv preprint arXiv:1806.01888*.

Belloni, Alexandre, Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. 2015. "Some new asymptotic theory for least squares series: Pointwise and uniform results." *Journal of Econometrics* 186 (2): 345–366. High Dimensional Problems in Econometrics.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2013. "Inference on Treatment Effects after Selection among High-Dimensional Controls." *The Review of Economic Studies* 81 (2): 608–650.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "Inference on treatment effects after selection among high-dimensional controls." *The Review of Economic Studies* 81 (2): 608–650.

Bianchi, Daniele. 2020. "Cryptocurrencies as an asset class? An empirical assessment." *The Journal of Alternative Investments* 23 (2): 162–179.

Bianchi, Daniele, and Mykola Babiak. 2021. "A risk-based explanation of cryptocurrency returns." *Available at SSRN 3935934*.

Bianchi, Daniele, Matthias Büchner, and Andrea Tamoni. 2021. "Bond risk premiums with machine learning." *The Review of Financial Studies* 34 (2): 1046–1089.

Bianchi, Daniele, Massimo Guidolin, and Manuela Pedio. 2022. "The dynamics of returns predictability in cryptocurrency markets." *The European Journal of Finance*: 1–29.

Bickel, Peter J., Ya'acov Ritov, and Alexandre B. Tsybakov. 2009. "Simultaneous analysis of Lasso and Dantzig selector." *The Annals of Statistics* 37 (4): 1705 – 1732.

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

Borri, Nicola, Daniele Massacci, Mirco Rubin, and Dario Ruzzi. 2022. "Crypto risk premia." *Available at SSRN*.

Bühlmann, Peter, and Sara van de Geer. 2011. *Statistics for high-dimensional data*. Springer Series in Statistics: xviii+556: Springer, Heidelberg. Methods, theory and applications.

Chamberlain, Gary. 1983. "Funds, factors, and diversification in arbitrage pricing models." *Econometrica: Journal of the Econometric Society*: 1305–1323.

Cheah, Jeremy Eng-Tuck, Di Luo, Zhuang Zhang, and Ming-Chien Sung. 2022. "Predictability of bitcoin returns." *The European Journal of Finance* 28 (1): 66–85.

Chen, Luyang, Markus Pelger, and Jason Zhu. 2020. "Deep learning in asset pricing." *Available at SSRN 3350138*.

Chen, Xiaohong. 2007. "Large Sample Sieve Estimation of Semi-Nonparametric Models." In *Handbook of Econometrics*, edited by J.J. Heckman and E.E. Leamer, vol. 6B 1st ed. chap. 76, 5549–5632: Elsevier.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. 2018a. "Double/debiased machine learning for treatment and structural parameters." *The Econometrics Journal*.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. 2018b. "Double/debiased machine learning for treatment and structural parameters." *The Econometrics Journal* 21 (1): C1–C68.

Chernozhukov, Victor, Christian Hansen, and Martin Spindler. 2015. "Valid post-selection and post-regularization inference: An elementary, general approach." *Annu. Rev. Econ.* 7

(1): 649–688.

Chetverikov, Denis, Zhipeng Liao, and Victor Chernozhukov. 2021. "On cross-validated Lasso in high dimensions." *The Annals of Statistics* 49 (3): 1300 – 1317.

Chetverikov, Denis, and Jesper Riis-Vestergaard Sorensen. 2021. "Analytic and bootstrap-after-cross-validation methods for selecting penalty parameters of high-dimensional M-estimators." *arXiv preprint arXiv:2104.04716*.

Chetverikov, Denis, and Jesper Riis-Vestergaard S$\omega rensen$.2021."$Analytic and Bootstrap - after - Cross - Validation Methods for Selecting Penalty Parameters of High - Dimensional M - Estimators."ArXivNA : 1 - -50$.

Cochrane, John H. 2009. *Asset Pricing.*: Princeton university press.

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

Cong, Lin William, George Andrew Karolyi, Ke Tang, and Weiyi Zhao. 2022. "Value premium, network adoption, and factor pricing of crypto assets."

Connor, Gregory, Matthias Hagmann, and Oliver Linton. 2012. "Efficient semiparametric estimation of the Fama–French model and extensions." *Econometrica* 80 (2): 713–754.

Connor, Gregory, and Oliver Linton. 2007. "Semiparametric estimation of a characteristic-based factor model of common stock returns." *Journal of Empirical Finance* 14 (5): 694–717.

Fama, Eugene F, and Kenneth R French. 1992. "The cross-section of expected stock returns." *the Journal of Finance* 47 (2): 427–465.

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

Fan, Jianqing, Yuan Liao, and Weichen Wang. 2016. "Projected principal component analysis in factor models." *Annals of statistics* 44 (1): 219.

Fan, Qingliang, Yu-Chin Hsu, Robert P Lieli, and Yichong Zhang. 2022. "Estimation of conditional average treatment effects with high-dimensional data." *Journal of Business & Economic Statistics* 40 (1): 313–327.

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu. 2020. "Taming the factor zoo: A test of new factors." *The Journal of Finance* 75 (3): 1327–1370.

Feng, Guanhao, Jingyu He, Nicholas G Polson, and Jianeng Xu. 2018. "Deep learning in characteristics-sorted factor models." *arXiv preprint arXiv:1805.01104*.

Fratzscher, Marcel. 2009. "What explains global exchange rate movements during the financial crisis?" *Journal of International Money and Finance* 28 (8): 1390–1407.

Giglio, Stefano, Bryan Kelly, and Dacheng Xiu. 2022. "Factor models, machine learning, and asset pricing." *Annual Review of Financial Economics* 14.

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

Giglio, Stefano, Dacheng Xiu, and Dake Zhang. 2021. "Test assets and weak factors."Technical report, National Bureau of Economic Research.

van der Greer, Sara. 2016. *Estimation and Testing under Sparsity*. Lecture Notes in Mathematics: Springer, New York, NY.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu. 2020. "Empirical asset pricing via machine learning." *The Review of Financial Studies* 33 (5): 2223–2273.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu. 2021. "Autoencoder asset pricing models." *Journal of Econometrics* 222 (1): 429–450.

Guijarro-Ordonez, Jorge, Markus Pelger, and Greg Zanotti. 2021. "Deep Learning Statistical Arbitrage." *Available at SSRN 3862004*.

Hayek, Friedrich A. 1976. *The Denationalization of Money: An Analysis of the Theory and Practice of Concurrent Currencies.*: The Institute of Economic Affairs.

Hu, Albert S, Christine A Parlour, and Uday Rajan. 2019. "Cryptocurrencies: Stylized facts on a new investible instrument." *Financial Management* 48 (4): 1049–1068.

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

Kelly, Bryan T, Seth Pruitt, and Yinan Su. 2020. "Instrumented principal component analysis." *Available at SSRN 2983919*.

Lamport, Leslie, Robert Shostak, and Marshall Pease. 1982. "The Byzantine generals problem." *ACM Transactions on Programming Languages and Systems* 4 (3): 382–401.

Lee, Jason D, Dennis L Sun, Yuekai Sun, and Jonathan E Taylor. 2016. "Exact post-selection inference, with application to the lasso."

Liebi, Luca J. 2022. "Is there a value premium in cryptoasset markets?" *Economic Modelling* 109: 105777.

Liu, Jiatao, Ian W Marsh, Paolo Mazza, and Mikael Petitjean. 2019. "Factor structure in cryptocurrency returns and volatility." *Available at SSRN 3389152*.

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

Liu, Yukun, Aleh Tsyvinski, and Xi Wu. 2021. "Accounting for cryptocurrency value." *Available at SSRN 3951514*.

Liu, Yukun, Aleh Tsyvinski, and Xi Wu. 2022. "Common risk factors in cryptocurrency." *The Journal of Finance* 77 (2): 1133–1177.

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

Nakamoto, Satoshi. 2008. "Bitcoin: A peer-to-peer electronic cash system."

Newey, Whitney. 1997. "Convergence rates and asymptotic normality for series estimators." *Journal of Econometrics* 79 (1): 147–168.

Newey, Whitney K., and Daniel McFadden. 1994. "Chapter 36 Large sample estimation and hypothesis testing." *Handbook of Econometrics* 4: 2111–2245.

Newey, Whitney K, and Kenneth D West. 1987. "Hypothesis testing with efficient method of moments estimation." *International Economic Review*: 777–787.

Perperoglou, Aris, Willi Sauerbrei, Michal Abrahamowicz, and Matthias Schmid. 2019. "A review of spline function procedures in R." *BMC medical research methodology* 19 (1): 1–16.

Lopez de Prado, Marcos. 2022. "Causal Factor Investing: Can Factor Investing Become Scientific?" *Available at SSRN 4205613*.

Ross, Stephen A. 1976. "The arbitrage theory of capital asset pricing." *Journal of economic theory* 13 (3): 341–360.

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66: 688–701.

Rubin, Donald B. 1978. "Bayesian Inference for Causal Effects." *The Annals of Statistics* 6 (1): 34–58.

Schmidt-Hieber, Johannes. 2020. "Nonparametric regression using deep neural networks with

ReLU activation function." *Annals of Statistics* 48: 1875–1897.

Semenova, Vira, and Victor Chernozhukov. 2021a. "Debiased machine learning of conditional average treatment effects and other causal functions." *The Econometrics Journal* 24 (2): 264–289.

Semenova, Vira, and Victor Chernozhukov. 2021b. "Debiased machine learning of conditional average treatment effects and other causal functions." *The Econometrics Journal* 24: 264–289. utaa027.

Shams, Amin. 2020. "The structure of cryptocurrency returns." *Fisher College of Business Working Paper* (2020-03): 011.

Shanken, Jay. 1992. "On the estimation of beta-pricing models." *The review of financial studies* 5 (1): 1–33.

Sharpe, William F. 1964. "The Capital Asset Pricing Model: A Theory of Market Equilibrium Under Conditions of Risk." *The Journal of Finance* 19 (3): 425–442.

Smucler, Ezequiel, Andrea Rotnitzky, and James M. Robins. 2019. "A unifying apptoach for doubly-robust ℓ_1 regularized estimation of causal contrasts." *ArXiv* NA: 1–125.

Tan, Zhiqiang. 2017. "Regularized calibrated estimation of propensity scores with model misspecification and high-dimensional data." *ArXiv* NA: 1–60.

Tan, Zhiqiang. 2020. "Model-assisted inference for treatment effects using regularized calibrated estimation with high-dimensional data." *The Annals of Statistics* 48 (2): 811 – 837.

Tibshirani, Robert. 1996. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society: Series B (Methodological)* 58 (1): 267–288.

Vershynin, Roman. 2018. *High-dimensional probability: An introduction with applications in data science*. vol. 47: Cambridge university press.

Wang, Wenjie, and Jun Yan. 2021. "Shape-Restricted Regression Splines with R Package splines2." *Journal of Data Science* 19 (3): 498–517.

Wu, Peng, Zhiqiang Tan, Wenjie Hu, and Xiao-Hua Zhou. 2021. "Model-Assisted Inference for Covariate-Specific Treatment Effects with High-dimensional Data."

Yao, Shouyu, Xiaoran Kong, Ahmet Sensoy, Erdinc Akyildirim, and Feiyang Cheng. 2021. "Investor attention and idiosyncratic risk in cryptocurrency markets." *The European Journal of Finance*: 1–19.

Yu, Yi, Tengyao Wang, and Richard J Samworth. 2015. "A useful variant of the Davis–Kahan theorem for statisticians." *Biometrika* 102 (2): 315–323.

Zhang, Wei, and Yi Li. 2020. "Is idiosyncratic volatility priced in cryptocurrency markets?" *Research in International Business and Finance* 54: 101252.

Zhang, Wei, and Yi Li. 2023. "Liquidity risk and expected cryptocurrency returns." *International Journal of Finance & Economics* 28 (1): 472–492.

Zhang, Wei, Yi Li, Xiong Xiong, and Pengfei Wang. 2021. "Downside risk and the cross-section of cryptocurrency returns." *Journal of Banking & Finance* 133: 106246.