Brief

Causality and Factor Investing: A Primer

Marcos López de Prado and Vincent Zoonekynd

CFA Institute Research Foundation

Causality and Factor Investing: A Primer

Marcos López de Prado
Vincent Zoonekynd

CFA Institute
Research
Foundation

Statement of Purpose

The CFA Institute Research Foundation is a not-for-profit organization established to promote the development and dissemination of relevant research for investment practitioners worldwide.

Photo credit: Getty Images / iStock / Getty Images Plus / SonerCdem

Ebook ISBN: 978-1-952927-59-1

CONTENTS

CFA Institute

PROFESSIONAL LEARNING QUALIFIED ACTIVITY
This publication qualifies for 0.5 PL credits under the guidelines of the CFA Institute Professional Learning Program.

CAUSALITY AND FACTOR INVESTING: A PRIMER

Marcos López de Prado
Global Head, Quantitative Research and Development, Abu Dhabi Investment Authority (ADIA); Board Member, ADIA Lab; Professor of Practice, College of Engineering, Cornell University; Research Fellow, Applied Mathematics and Computational Research Department, Lawrence Berkeley National Laboratory

Vincent Zoonekynd
Lead, Quantitative Research and Development, Abu Dhabi Investment Authority (ADIA); Research Affiliate, ADIA Lab

Introduction

Factor investing was once heralded as the future of systematic asset management. Academic breakthroughs such as Fama and French's (1993) three-factor model gave rise to a burgeoning field, with hundreds of anomalies proposed and institutionalized (Harvey, Liu, and Zhu 2016). Multifactor products now accumulate trillions of dollars in assets under management, promising style tilts that enhance returns, reduce risk, or both. Yet the real-world long-term performance of most such strategies has fallen short of expectations (López de Prado and Zoonekynd 2026).

This shortfall has prompted soul-searching across the industry. Critics cite *p*-hacking and backtest overfitting as the main culprits (see Bailey, Borwein, López de Prado, and Zhu 2014; Fabozzi and López de Prado 2018; and Harvey and Liu 2020). Others argue that market participants arbitrage these opportunities shortly after publication (McLean and Pontiff 2016). Still others suggest that factors work only in certain regimes and that the recent regime has been unfavorable—a dubious *ex post* argument given that the original publications made no regime distinctions (Evans 1994; Anderson 2011).

These explanations may contain elements of truth, but they overlook a deeper problem: Association does not imply causation. The econometric canon applied in factor investing studies—linear regression, two-pass estimation, *p*-values, and correlation-based statistics—rarely discusses causality. Yet investment decisions are inherently causal, because they require the attribution of returns to risk sources. We do not merely want to know if a portfolio composed of high book-to-market stocks delivers positive returns. To optimize a portfolio, we need to determine how much of that performance is attributable to (i.e., caused by) the value factor, to the exclusion of (i.e., controlling for) other explanations, and identify what may disrupt this relationship (i.e., causal mechanism) in the future.

The views expressed in this paper are the authors' and do not necessarily represent the opinions of the organizations with which they are affiliated. We thank our ADIA colleagues, especially Pascal Blanqué, Alexander Lipton, Anders Svennesen, and Jean-Paul Villain, for their suggestions. The article has also benefited from conversations with Patrick Cheridito (ETH Zürich), Gerald Cubbin (Invesco), Robert F. Engle (NYU), Frank Fabozzi (EDHEC), Luis Garcia-Feijoo (Florida Atlantic University), Campbell Harvey (Duke University), Miguel Hernán (Harvard University), Guido Imbens (Stanford University), Alessia López de Prado Rehder (University of Zurich), Riccardo Rebonato (EDHEC), Alessio Sancetta (Royal Holloway University of London), Luis Seco (University of Toronto), Horst Simon (ADIA Lab), and Josef Teichmann (ETH Zürich).

The distinction between association and causation goes beyond semantics. Some associational models may produce good forecasts without offering a risk–return attribution, thus exposing investors to unknown or unwanted risks. Without accounting for causal structure, models are likely to be biased, unstable, and unprofitable when run out of sample.[1]

From "Factor Zoo" to "Factor Mirage"

Cochrane (2011) coined the term "factor zoo" to illustrate the explosion of empirical findings in asset pricing. Hundreds of published anomalies now compete for attention, yet most fail to survive replication or implementation. The literature has responded with tools to mitigate data snooping, such as corrections for multiple testing, the deflated Sharpe ratio (DSR), Bayesian shrinkage, and out-of-sample tests (Bailey and López de Prado 2014; López de Prado 2018, 2020).

Although useful, these techniques do not address inference errors due to model misspecification. A model can be *p*-hacking-free and still misspecified, leading to false positives or false negatives. In this article, we introduce the concept of the "factor mirage": an empirical finding that appears sound by conventional statistical standards but is structurally invalid because it misrepresents the causal relationships among variables, leading to a biased risk and return attribution.

Factor mirages arise from two common specification errors: (1) confounder bias—failing to control for variables that are *causes* of both an independent variable (factor) and the dependent variable (returns); and (2) collider bias—controlling for variables that are *consequences* of both an independent variable and the dependent variable. In econometric terms, these biases introduce noncausal associations, distorting coefficient estimates. In financial terms, they lead to misattributed risks and returns, inefficient investments, and even systematic losses.

Unlike brute-force *p*-hacking, which is an acknowledged malpractice, the factor mirage is subtler. It arises from practices that are widely taught, widely applied, and rarely questioned.

Where the Canon Fails: Econometrics Without Causality

The econometric methods most widely used in empirical finance—ordinary least squares regressions, stepwise model selection, and significance testing—assume that the regression model is correctly specified and that residuals are stationary and behave like white noise. These assumptions are often violated in financial applications, particularly in asset pricing models.

---

[1]Although causality has a long history in economics going back to Smith (1776), the term is often used with different and confusing meanings. For example, Haavelmo (1944) discussed causality in the *ceteris paribus* sense, which simulates a controlled experiment. In contrast, Granger (1969, 1980) discussed causality in a predictive (associational) sense. For Granger, a variable X causes Y if X has predictive power over Y after controlling for lags of Y. This dynamic is not strict causality, because it does not contemplate counterfactuals (Pearl 2009; Pearl, Glymour, and Jewell 2016). For example, Granger's approach will find that a rooster causes the sun to rise, whereas a controlled experiment will show that the sun will also rise in the absence of a rooster. Modern economists working on (counterfactual) causality include Guido Imbens, Joshua Angrist, and David Card. Nonetheless, these concepts have not yet permeated financial economics, in which associational statistics remains ubiquitous.

A case in point is the assumption of correct model specification. In asset pricing, the standard practice for identifying a factor is to follow a two-pass regression approach: First, run time-series regressions of assets' excess returns on a set of factors to estimate factor loadings (exposures). Second, run cross-sectional regressions of assets' excess returns on the estimated factor exposures to estimate factor premia. The choice of specification is typically driven by associational power maximization, not causal considerations.

A confounder is a variable that is a cause of both an explanatory variable and the dependent variable. Confounder bias arises when the model's specification does not control for a confounder.[2] If leverage, for instance, influences both the book-to-market ratio and returns, and leverage is not included in the model, the estimated coefficient for book-to-market ratio will be biased in magnitude and perhaps sign. Confounder bias exposes investors to unwanted risks and premia.

A collider is a variable that is causally downstream of both an explanatory variable and the dependent variable. Controlling for a collider introduces a noncausal correlation, which biases coefficient estimates, inflates the adjusted R^2, and tends to lower p-values. For example, if quality is influenced by both book-to-market and returns, including it as a control will bias the estimated coefficient for book-to-market. This collider bias is subtle: It does not cause multicollinearity, and it reduces standard errors, but it distorts inference. Two aspects make colliders particularly dangerous for investors: First, they can change the sign of estimated coefficients, thus inducing investors to buy securities that should be sold and to sell securities that should be bought;[3] and, second, the noncausal association created by a collider cannot be monetized. By the time the collider is observed, the value of the dependent variable is already set. The effect estimated in the regression and the performance simulated in the backtest are a mirage.[4] See exhibit 5 for numerous examples.

The standard two-pass procedure is particularly vulnerable to collider bias. Researchers' inclusion or exclusion of controls is typically justified on statistical criteria (e.g., increasing R^2) rather than causal logic; see Fama and French (1993, 2015). Similarly, the three-pass factor regression approach in Giglio and Xiu (2021) applies principal component analysis (PCA) to the matrix of asset returns to extract latent variables—that is, directions of maximal variance in returns unexplained by observed factors—assuming without causal evidence that those latent variables must be confounders. The problem is that PCA attempts to maximize explained variance without differentiating between confounders and colliders. As a result, model specifications that look compelling when in-sample will often introduce strong collider biases in the estimated coefficients.

Standard model evaluation metrics—such as adjusted R^2, Akaike information criterion, Bayesian information criterion, and t-statistics—reward misspecification and penalize parsimony, even when the extra variables introduce collider bias. In a world of limited data and noisy signals, these practices create an illusion of robustness and profitability. It is also important to recognize

[2]The econometrics literature sometimes refers to this concept as "omitted variable bias."

[3]These concerns are not hypothetical. Shanken (1992) discussed the consequences of estimation error in factor betas. Giglio and Xiu (2021) showed that many popular factors are likely mispriced because of omitted variables and introduced an intermediate principal component analysis step to identify potential latent confounders. Although these considerations partially address concerns regarding confounder bias, the problem of collider bias has received virtually no attention in the finance literature.

[4]For a review of backtesting methods, see Joubert, Sestovic et al. (2024).

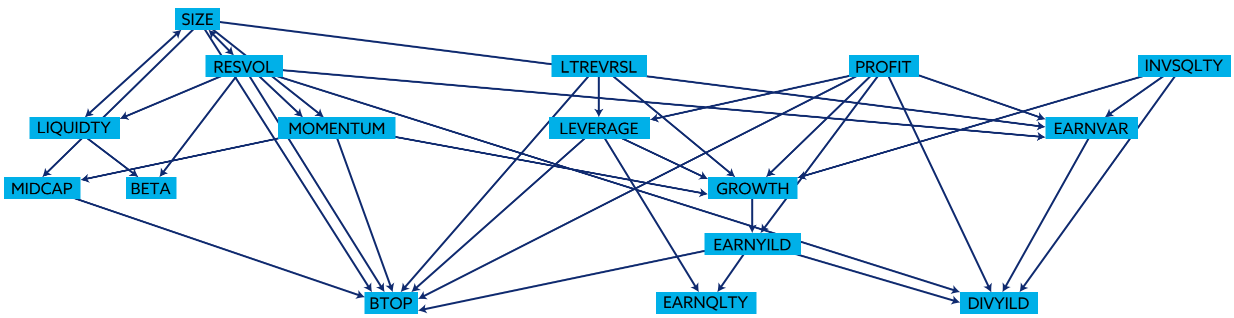
that multiple testing adjustments do not correct for model misspecification: A model can be misspecified after a single test.

The essence of the factor mirage is a model that is methodologically correct by current econometric standards but that fails to capture the underlying causal structure. It performs well in backtests and cross-validation but delivers disappointing results out of sample and in live trading, because noncausal associations are not monetized.

Example: Colliders Among Barra Factors

To illustrate these points, we apply the Peter-Clark (PC) causal discovery algorithm of Spirtes, Glymour, and Scheines (2000) to the time series of daily returns for the risk factors of 85 Barra risk models. This algorithm finds the network of dependencies among variables, which are visualized in the form of causal graphs (one causal graph for each model).[5] **Exhibit 1** aggregates the resulting causal graphs by retaining the edges present in at least one-third of the graphs.

Exhibit 1. Aggregate Causal Graph Discovered through the PC Algorithm



Notes: The Barra risk factors are defined from the following stock-level characteristics: BETA (sensitivity to the cap-weighted excess returns of the estimation universe); BETANL: non-linear beta (cube of BETA, orthogonalized); BTOP: book-to-price ratio; DIVYILD: dividend yield (dividend-to-price ratio); EARNQLTY: earnings quality (accruals); EARNVAR: earnings variability (standard deviation of earnings, sales, or cash flows divided by the average); EARNYILD: earnings yield (average of predicted and trailing earnings-to-price ratios); GROWTH (average of long-term predicted earnings growth and past 5-year earnings growth); INVSQLTY: investment quality (average of total assets growth rate, issuance growth, capital expenditures growth); LEVERAGE (weighted average of leverage, defined as (value of common equity + preferred equity + long-term debt) / common equity, and debt-to-assets ratio); LIQUIDTY (monthly share turnover); LTREVRSL: long-term (5-year) reversal; MIDCAP: mid capitalization (cube of the SIZE factor, orthogonalized to SIZE); MOMENTUM (2-year momentum, excluding the latest month); PROFIT: profitability (average of asset turnover, profitability, profit margin, return on assets); RESVOL: residual volatility; SIZE (log of market cap); SIZE (log or market cap); SIZENL: non-linear size (cube of SIZE, orthogonalized).

Source: Barra.

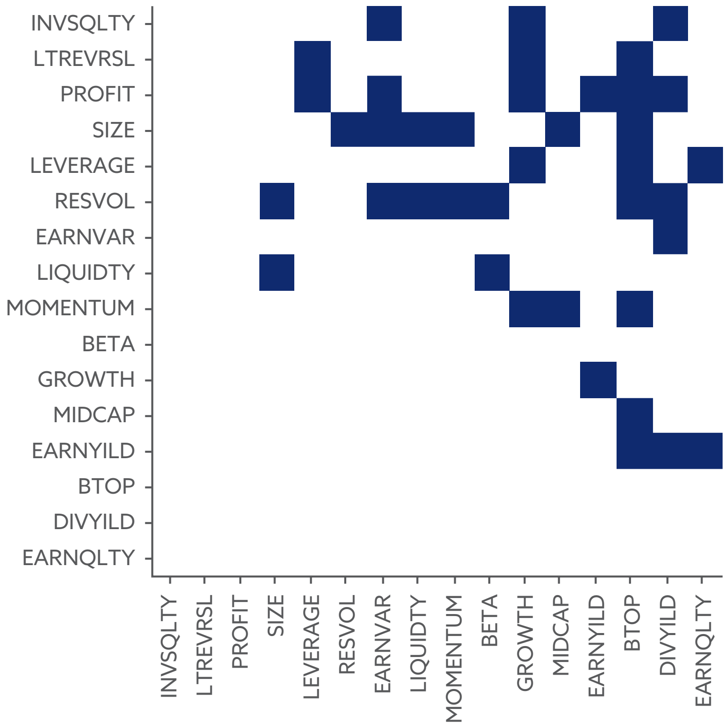
[5]A causal graph is a mathematical graph that connects two variables when one is a function of the other. The arrow starts at the cause and points to the variable that is dependent (i.e., where the effect is observed). Causal graphs allow us to visualize the structure of dependencies in a system and identify the correct controls in an experiment. For a discussion of causal discovery algorithms, see Glymour, Zhang, and Spirtes (2019) and Olivetti et al. (forthcoming 2025).

Exhibit 2 derives the corresponding adjacency matrix. The two entries below the main diagonal in this exhibit indicate the edges for which the PC algorithm could not determine the direction of the causal relationship, namely (SIZE, LIQUIDITY) and (SIZE, RESVOL).

With the help of this discovered graph, an investor can formulate the correct model specification. In particular, to avoid the risk of controlling for a collider, an investor aiming to invest in one of those factors should condition on the ancestors of that factor, not on its descendants. For instance, to invest on "growth," we should control for "momentum," "leverage," "long-term reversal," "profit," and "investor quality" but not for value factors (e.g., "earnings yield," "book-to-price," "earnings quality," "dividend yield"). **Exhibit 3** highlights in blue the correct controls, in yellow the controls to be avoided, and in purple the irrelevant variables.

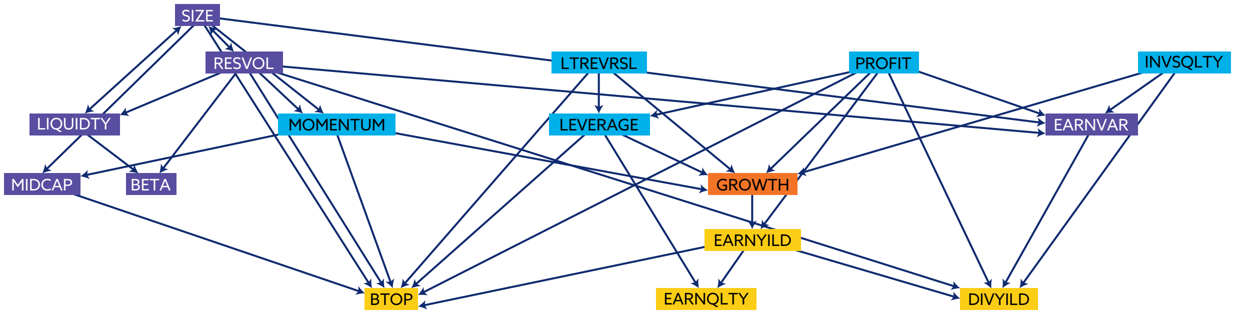
If we take one of those risk models, say the Barra US equity model USE4L, we can forecast "growth" from the other factors, selecting them to maximize the adjusted R^2. The left plot in **Exhibit 4** shows the adjusted R^2 of models that include all descendants, in which a greedy algorithm adds ancestors (in the *y*-axis) one by one. The full model, with all descendants and the ancestors listed in the *y*-axis, has an adjusted R^2 of approximately 8.5%. These models are misspecified, because descendants should not have been included as control variables. The right plot in Exhibit 4 shows models without descendants, where a greedy algorithm adds ancestors (in the *y*-axis) one by one. The full model, with all ancestors but no descendants, has an adjusted R^2 of approximately 7.8%, which is "worse" than the adjusted R^2 of the misspecified model.

Exhibit 2. Adjacency Matrix



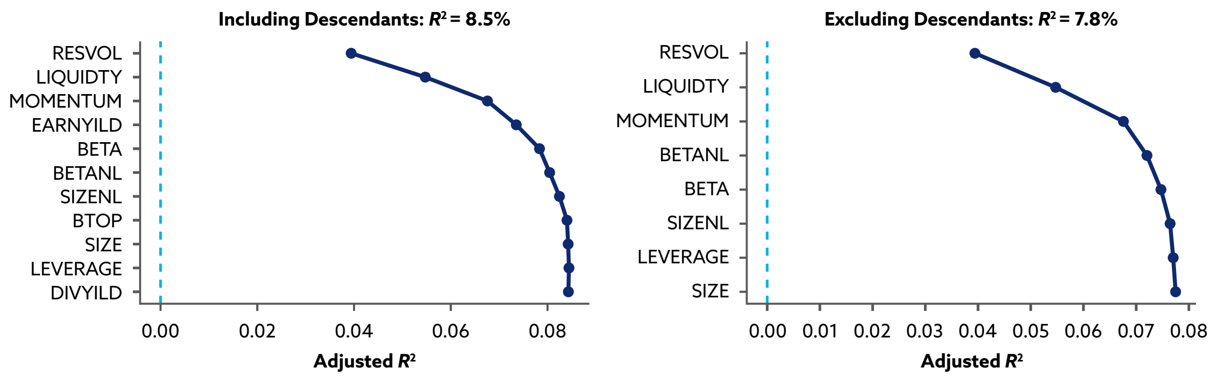
Source: Barra.

Exhibit 3. Correct and Incorrect Controls of a Growth Factor Model



Source: Barra.

Exhibit 4. Adjusted R^2 for Correctly (right) and Incorrectly (left) Specified Factor Models



Source: Barra.

Exhibit 5 shows 26 factor models for which adding a collider changes the sign of the estimated coefficient. The "Factor" column shows the explained risk factor, the "Cause" column shows a correct control, and the "Correct Beta" and "Correct aR²" columns show the estimated coefficient and adjusted R^2 when the model applies the correct control. Adding as a control the variable listed under the "Collider" column increases the adjusted R^2, as shown in the "Biased aR²" column, but it also changes the sign on the estimated coefficient, as shown in the "Biased Beta" column. This is an evolving, partial list and is presented for illustrative purposes only.

This finding confirms the danger of using three-pass regression approaches and black-box machine learning models. Any choice of controls should always be argued through a causal graph and supported by empirical evidence that those controls do not include a collider. Furthermore, the prevalence of financial colliders also argues against the "Virtue of Complexity" approach (i.e., the notion that massively overfit financial models perform better than the slightly overfit ones; see Berk 2023; Buncic 2025; Cartea, Jin, and Shi 2025;

Exhibit 5. Examples of Barra Risk Models: Adding a Collider Changes the Sign of the Estimated Exposures

Model	Factor	Cause	Correct Beta	Correct aR^2	Collider	Biased Beta	Biased aR^2
INE2L	LIQUIDTY	RESVOL	0.0824	0.0070	BETA	−0.0362	0.0769
INE2S	LIQUIDTY	RESVOL	0.0824	0.0070	BETA	−0.0362	0.0769
THE2L	LIQUIDTY	RESVOL	0.0110	0.0000	BETA	−0.0231	0.0085
THE2S	LIQUIDTY	RESVOL	0.0110	0.0000	BETA	−0.0231	0.0085
CLE2L	EARNYILD	GROWTH	0.0817	0.0964	EARNQLTY	−0.0280	0.2316
CLE2S	EARNYILD	GROWTH	0.0817	0.0964	EARNQLTY	−0.0280	0.2316
OME2L	MIDCAP	MOMENTUM	0.0159	0.0005	BTOP	−0.0014	0.0112
OME2S	MIDCAP	MOMENTUM	0.0159	0.0005	BTOP	−0.0014	0.0112
IDE2L	EARNVAR	PROFIT	−0.0050	0.0777	DIVYILD	0.0129	0.0930
IDE2S	EARNVAR	PROFIT	−0.0050	0.0777	DIVYILD	0.0129	0.0930
MLE2L	LEVERAGE	LTREVRSL	0.0056	0.0088	EARNQLTY	−0.0014	0.0288
MLE2S	LEVERAGE	LTREVRSL	0.0056	0.0088	EARNQLTY	−0.0014	0.0288
NGE2L	LEVERAGE	LTREVRSL	0.0125	0.0151	BTOP	−0.0008	0.0199
NGE2S	LEVERAGE	LTREVRSL	0.0125	0.0151	BTOP	−0.0008	0.0199
AEE2L	GROWTH	MOMENTUM	0.0086	0.0326	BTOP	−0.0006	0.0502
AEE2S	GROWTH	MOMENTUM	0.0086	0.0326	BTOP	−0.0006	0.0502
EMM1L	GROWTH	MOMENTUM	0.0003	0.0006	BTOP	−0.0003	0.0036
EMM1S	GROWTH	MOMENTUM	0.0003	0.0006	BTOP	−0.0003	0.0036
AEE2L	MOMENTUM	RESVOL	−0.0064	−0.0002	MIDCAP	0.0017	0.0273
AEE2L	MOMENTUM	RESVOL	−0.0064	−0.0002	DIVYILD	0.0019	0.0020
AEE2S	MOMENTUM	RESVOL	−0.0064	−0.0002	DIVYILD	0.0019	0.0020
AEE2S	MOMENTUM	RESVOL	−0.0064	−0.0002	MIDCAP	0.0017	0.0273
MXE2L	GROWTH	LEVERAGE	−0.0021	0.0238	EARNQLTY	0.0032	0.0242
MXE2S	GROWTH	LEVERAGE	−0.0021	0.0238	EARNQLTY	0.0032	0.0242
USSLOWL	GROWTH	LTREVRSL	0.0022	0.0589	DIVYILD	−0.0009	0.0804
USSLOWS	GROWTH	LTREVRSL	0.0022	0.0589	DIVYILD	−0.0009	0.0804

Note: Most Barra model codes have a 2-letter country code (AE: UAE, CL: Chile, ID: Indonesia, IN: India, ML: Malaysia, MX: Mexico, NG: Nigeria, OM: Oman, TH: Thailand), followed by E for equity, a version number, and L (long-term) or S (short-term). There are two exceptions in this table: EMM (Emerging markets) and USSLOW (US total market long-term).

Source: Barra.

Fallahgoul 2025; and Nagel 2025). In other words, given the ubiquity of colliders in financial datasets, it is generally better to underfit than overfit—if you must err, err conservatively. A slightly underfit model may still generalize reasonably, its backtested performance is less likely to be a mirage, and it will be more robust to parameter shifts and structural breaks.

The Economic Cost of Causal Neglect

Misapplying associational tools to causal problems is not only a technical misstep; it has real-world consequences. Portfolios are constructed, risks are hedged, and billions of dollars are deployed based on models that are misspecified.

The economic costs of causal neglect fall into several categories:

- *Capital misallocation*. Investors allocate to strategies that appear statistically significant but are not economically meaningful. This misdirection persists until performance disappoints or capital is withdrawn (López de Prado, Lipton, and Zoonekynd forthcoming 2025).
- *Hidden leverage and risk stacking*. When multiple models share similar specification errors, portfolios may unknowingly stack risk exposures. For example, many value strategies may be exposed to the same macroeconomic confounders (López de Prado 2023).
- *Excessive turnover*. Spurious signals lead to unnecessary trades, increasing transaction costs, bid–ask spreads, and slippage. This turnover further erodes alpha, to the benefit of competitors.
- *Lack of persistence*. Models built on noncausal relationships often fail to persist when economic conditions change or new data become available. A shift in a collider's parameter can spuriously flip the sign of the estimated risk premium. This shift explains the so-called time-varying risk premia conundrum. Investment models based on causal relationships can be profitable even if parameters shift (López de Prado and Zoonekynd forthcoming 2026).
- *Loss of trust*. When backtests consistently outperform live performance, clients lose trust in academic work and the scientific validity of systematic investing. This reputational damage affects academics and practitioners alike, and it is difficult to reverse (López de Prado 2015).

In short, causal neglect leads to inefficient investment decisions, unrewarded risks, and poor stewardship of capital. The current state of factor investing—marked by underwhelming performance, crowded trades, and skepticism—can be traced in part to these methodological shortcomings.

Best Practices for Professionals and Asset Owners

We propose the following checklist that practitioners and managers can use to assess whether a factor investing proposal is supported by causal evidence. These questions can form part of a due diligence questionnaire, investment memo, or strategy approval process. For a detailed description of these steps, see López de Prado and Zoonekynd (forthcoming 2026).

Step 1: Variable Selection

- What is the intended purpose of the factor model: risk attribution or risk premia harvesting? Are the selected variables consistent with this purpose?
- How were the candidate variables initially selected?
- Were nonparametric or machine learning methods used to detect relationships?
- Were Shapley values, mean decrease impurity, or feature importance used?
- Were domain-specific constraints applied to exclude spurious or uninterpretable variables?

Step 2: Causal Discovery

- Did the researcher construct a causal graph to represent the structure of the problem?
- Were causal discovery algorithms used? If so, which ones (e.g., Peter-Clark (PC), linear non-Gaussian acyclic model (LiNGAM), greedy equivalence search (GES))?
- What economic rationale or domain expertise supports the chosen graph?
- Have alternative causal graphs been considered and ruled out? On what basis?
- Are the causal graph's assumptions clearly documented and available for review?
- Has the causal graph structure been tested, using Microsoft's DoWhy "refutations" functionality?

Step 3: Causal Adjustment

- What method was used to identify the adjustment (e.g., backdoor, front door, instrumental variable)?
- Which variables are being controlled for, and why?
- Did researchers confirm the validity of the adjustment using do-calculus software (e.g., DAGitty,[6] DoWhy, or networkX's is_d_separator function)?
- Were any known colliders included in the model? How were they identified and excluded?
- Are all control variables economically interpretable and justifiable?

Step 4: Causal Explanatory and Predictive Power

- Does the model allow for the simulation of controlled experiments? Does it answer counter-factual questions?
- How was the model's generalization error estimated? Is this approach realistic, given the causal graph?
- Was the model's performance assessed in terms of: (1) probability of return sign, (2) ranking of returns, and (3) magnitude of returns?
- Were the findings robust across multiple validation techniques or subsamples?

---

[6]See the DAGitty website, www.dagitty.net.

- Does the model show signs of overfitting to any specific performance metric?
- What is the strength and robustness of the causal effect? How strong should a missing confounder be to change the sign of the effect? Does removing a supposed confounder change the sign of the effect?

Step 5: Causal Portfolio Construction

- How are causal effects translated into portfolio weights?
- Does the strategy remain neutral or agnostic to noncausal factors?
- Is the causal graph used to hedge unwanted exposures or to guide hedging?
- Are stress tests performed on the causal model (e.g., directed acyclic graph perturbation)?
- What portfolio constraints and transaction costs are considered, and how are they incorporated?
- Is the transfer coefficient between the ideal and actual portfolio computed and reported?

Step 6: Backtesting Methodology

- Which backtest methods were used (e.g., walk-forward, resampling, Monte Carlo)?
- Was the causal graph used to simulate scenarios?
- Are limitations of historical resampling or burn-in periods acknowledged?
- Are multiple scenarios or paths evaluated, rather than only one historical realization?
- Was combinatorially purged cross-validation or a similar technique used to estimate a distribution of Sharpe ratios?

Step 7: Multiple Testing Adjustment

- Are all models or hypotheses tested during research available to peer reviewers?
- Was the effective number of tests estimated, accounting for test correlation?
- Were *p*-value adjustments (e.g., Holm, Hochberg, Benjamini–Hochberg) applied?
- Was the deflated Sharpe ratio (DSR) reported? If so, how was it computed?
- Are all statistical significance levels adjusted for the model selection process?

Final Assessment

- Is the entire research process transparent, documented, and reproducible?
- Are assumptions stated explicitly and subject to falsifiability or peer review?
- Has the model been stress-tested against changes in its structure or estimation method?
- Does the investment team demonstrate familiarity with causal inference concepts and limitations?

By insisting on these standards, asset owners and supervisors can protect capital, enhance accountability, and improve alignment between strategy design and economic intent.

Conclusion

Investment factors exist, but the way researchers build and evaluate them is flawed. The econometric canon's neglect of causality has led to a proliferation of anomalies, most of which fail to hold up under scrutiny or deliver in practice. This phenomenon, which we call the factor mirage, reflects the consequences of model misspecification—particularly collider bias and confounder bias—within canonical estimation frameworks.

Our goal is not to impose a rigid orthodoxy but rather to call for better causal reasoning. Despite a long tradition in economics, causal reasoning has largely been ignored in finance. Fields such as medicine, policy evaluation, and macroeconomics have already embraced causal inference with measurable benefits. It is time for asset management to evolve.

The transition from associational to causal factor modeling will not be easy. It will require rewriting textbooks, questioning accepted dogmas, upgrading academic programs, unlearning old habits, revisiting familiar practices, and retooling teams. But the rewards are worth it: more stable strategies, better risk control, clearer communication, and ultimately more trustworthy products.

In a world increasingly skeptical of backtests and statistical alchemy, causal factor investing offers a credible path forward. It replaces the illusion of precision with the discipline of structure, helping us see beyond the mirage.

REFERENCES

Anderson, Robert M. 2011. "Time-Varying Risk Premia." *Journal of Mathematical Economics* 47 (3): 253–59. doi:10.1016/j.jmateco.2010.12.010.

Bailey, David H., Jonathan M. Borwein, Marcos López de Prado, and Qiji Jim Zhu. 2014. "Pseudo-Mathematics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-of-Sample Performance." *Notices of the American Mathematical Society* 61 (5): 458–71. doi:10.1090/noti1105.

Bailey, David H., and Marcos López de Prado. 2014. "The Deflated Sharpe Ratio: Correcting for Selection Bias, Backtest Overfitting and Non-Normality." *Journal of Portfolio Management* 40 (5): 94–107. doi:10.3905/jpm.2014.40.5.094.

Berk, Jonathan. 2023. "Comment on 'The Virtue of Complexity in Return Prediction'." doi:10.2139/ssrn.4410125.

Buncic, Daniel. 2025. "Simplified: A Closer Look at the Virtue of Complexity in Return Prediction." Working paper, SSRN. doi:10.2139/ssrn.5239006.

Cartea, Álvaro, Qi Jin, and Yuantao Shi. 2025. "The Limited Virtue of Complexity in a Noisy World." Working paper, SSRN. doi:10.2139/ssrn.5202064.

Cochrane, John H. 2011. "Presidential Address: Discount Rates." *Journal of Finance* 66 (4): 1047–1108. doi:10.1111/j.1540-6261.2011.01671.x.

Evans, Martin D. 1994. "Expected Returns, Time-Varying Risk, and Risk Premia." *Journal of Finance* 49 (2): 655–79. doi:10.1111/j.1540-6261.1994.tb05156.x.

Fabozzi, Frank, and Marcos López de Prado. 2018. "Being Honest in Backtest Reporting: A Template for Disclosing Multiple Tests." *Journal of Portfolio Management* 45 (1): 141–47. doi:10.3905/jpm.2018.45.1.141.

Fallahgoul, Hasan. 2025. "High-Dimensional Learning in Finance." Working paper (4 June). doi:10.48550/arXiv.2506.03780.

Fama, Eugene F., and Kenneth R. French. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics* 33 (1): 3–56. doi:10.1016/0304-405X(93)90023-5.

Fama, Eugene F., and Kenneth R. French. 2015. "A Five-Factor Asset Pricing Model." *Journal of Financial Economics* 116 (1): 1–22. doi:10.1016/j.jfineco.2014.10.010.

Giglio, Stefano, and Dacheng Xiu. 2021. "Asset Pricing with Omitted Factors." *Journal of Political Economy* 129 (7): 1947–90. doi:10.1086/714090.

Glymour, Clark, Kun Zhang, and Peter Spirtes. 2019. "Review of Causal Discovery Methods Based on Graphical Models." *Frontiers in Genetics* 10. doi:10.3389/fgene.2019.00524.

Granger, C. W. J. 1969. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica* 37 (3): 424–38. doi:10.2307/1912791.

Granger, C. W. J. 1980. “Testing for Causality: A Personal Viewpoint.” *Journal of Economic Dynamics and Control* 2:329–52. doi:10.1016/0165-1889(80)90069-X.

Haavelmo, Trygve. 1944. “The Probability Approach in Econometrics.” *Econometrica* 12 (Supplement): 1–115. doi:10.2307/1906935.

Harvey, Campbell R., and Yan Liu. 2020. “False (and Missed) Discoveries in Financial Economics.” *Journal of Finance* 75 (5): 2503–53. doi:10.1111/jofi.12951.

Harvey, Campbell R., Yan Liu, and Heqing Zhu. 2016. “. . . and the Cross-Section of Expected Returns.” *Review of Financial Studies* 29 (1): 5–68. doi:10.1093/rfs/hhv059.

Joubert, Jacques, Dragan Sestovic, Illya Barziy, Walter Distaso, and Marcos López de Prado. 2024. “Enhanced Backtesting for Practitioners.” *Journal of Portfolio Management* 51 (2): 12–27. doi:10.3905/jpm.2024.1.637.

López de Prado, Marcos. 2015. “The Future of Empirical Finance.” *Journal of Portfolio Management* 41 (4): 140–44. doi:10.3905/jpm.2015.41.4.140.

López de Prado, Marcos. 2018. *Advances in Financial Machine Learning*. New York: John Wiley & Sons.

López de Prado, Marcos. 2020. *Machine Learning for Asset Managers*. Cambridge, UK: Cambridge University Press. doi:10.1017/9781108883658.

López de Prado, Marcos. 2023. *Causal Factor Investing: Can Factor Investing Become Scientific?* Cambridge, UK: Cambridge University Press. doi:10.1017/9781009397315.

López de Prado, Marcos, and Vincent Zoonekynd. Forthcoming 2026. “Correcting the Factor Mirage: A Research Protocol for Causal Factor Investing.” *Journal of Portfolio Management*. Working paper, FEB-RN Research Paper No. 101/2025, titled “Why Has Factor Investing Failed?: The Role of Specification Errors,” available at doi:10.2139/ssrn.4697929.

López de Prado, Marcos, Alexander Lipton, and Vincent Zoonekynd. Forthcoming 2025. “Causal Factor Analysis Is a Necessary Condition for Investment Efficiency.” *Journal of Portfolio Management*. Working paper available at doi:10.2139/ssrn.5131050.

McLean, R. David, and Jeffrey Pontiff. 2016. “Does Academic Research Destroy Stock Return Predictability?” *Journal of Finance* 71 (1): 5–32. doi:10.1111/jofi.12365.

Nagel, Stefan. 2025. “Seemingly Virtuous Complexity in Return Prediction.” Chicago Booth Research Paper No. 25–10. doi:10.2139/ssrn.5335012.

Olivetti, Emanuele, Vincent Zoonekynd, Patrick Yam, Marcos López de Prado, Guido Imbens, and Miguel Hernán. Forthcoming 2025. “ADIA Lab Causal Discovery Challenge.” *Journal of Financial Data Science*.

Pearl, Judea. 2009. *Causality: Models*. 2nd ed. Cambridge University Press. doi:10.1017/CBO9780511803161.

Pearl, Judea, Madelyn Glymour, and Nicholas P. Jewell. 2016. *Causal Inference in Statistics: A Primer*. New York: Wiley.

Shanken, Jay. 1992. "On the Estimation of Beta-Pricing Models." *The Review of Financial Studies* 5 (1): 1–33.

Smith, Adam. (1776) 1976. *The Wealth of Nations*. Edited by R. H. Campbell and A. S. Skinner. Glasgow Edition of The Works and Correspondence of Adam Smith, Vol. 2a.

Spirtes, Peter, Clark Glymour, and R. Scheines. 2000. *Causation, Prediction, and Search*. 2nd ed. MIT Press.

Named Endowments

CFA Institute Research Foundation acknowledges with sincere gratitude the generous contributions of the Named Endowment participants listed below.

Gifts of at least US$100,000 qualify donors for membership in the Named Endowment category, which recognizes in perpetuity the commitment toward unbiased, practitioner-oriented, relevant research that these firms and individuals have expressed through their generous support of CFA Institute Research Foundation.

Ameritech
Anonymous
Robert D. Arnott
Theodore R. Aronson, CFA
Asahi Mutual Life Insurance Company
Batterymarch Financial Management
Boston Company
Boston Partners Asset Management, L.P.
Gary P. Brinson, CFA
Brinson Partners, Inc.
Capital Group International, Inc.
Concord Capital Management
Dai-Ichi Life Insurance Company
Daiwa Securities
Mr. and Mrs. Jeffrey Diermeier
Gifford Fong Associates
John A. Gunn, CFA
Investment Counsel Association of America, Inc.
Jacobs Levy Equity Management
Jon L. Hagler Foundation
Long-Term Credit Bank of Japan, Ltd.
Lynch, Jones & Ryan, LLC
Meiji Mutual Life Insurance Company
Miller Anderson & Sherrerd, LLP
John B. Neff, CFA
Nikko Securities Co., Ltd.
Nippon Life Insurance Company of Japan
Nomura Securities Co., Ltd.
Payden & Rygel
Provident National Bank
Frank K. Reilly, CFA
Salomon Brothers
Sassoon Holdings Pte. Ltd.
Scudder Stevens & Clark
Security Analysts Association of Japan
Shaw Data Securities, Inc.
Sit Investment Associates, Inc.
Standish, Ayer & Wood, Inc.
State Farm Insurance Company
Sumitomo Life America, Inc.
T. Rowe Price Associates, Inc.
Templeton Investment Counsel Inc.
Frank Trainer, CFA
Travelers Insurance Co.
USF&G Companies
Yamaichi Securities Co., Ltd.

Senior Research Fellows

Financial Services Analyst Association

For more on upcoming CFA Institute Research Foundation publications and webcasts, please visit www.cfainstitute.org/research/foundation.

Available online at rpc.cfainstitute.org