The Review of Asset Pricing Studies

An Empirical Assessment of Characteristics and Optimal Portfolios

Christopher G. Lamoureux
Eller College of Management, The University of Arizona, USA

Huacheng Zhang
University of Edinburgh Business School, UK

We implement a dynamically regularized, bootstrapped two-stage out-of-sample parametric portfolio policy to evaluate characteristics' efficacy in the conditional stock return-generating process in the metric of expected power utility. Traditional characteristics, such as momentum and size afforded large utility gains before 1999. These opportunities have since vanished. Overfitting—imprecision in weight estimation—is correlated with the optimal portfolio's variance. Therefore, it is not a problem for power utility investors with coefficients of relative aversion greater than four. For more risk-tolerant investors, we successfully reduce estimation error by increasing the curvature of the loss function relative to the investor's utility function. (*JEL* L200; C110; C350)

Received: 10 June 2022; Editorial Decision: 1 January 2024
Editor: Clemens Sialm

Much of the empirical research in asset pricing over the past 40 years examines the predictive content of measurable stock characteristics. We are interested in the question of whether risk-averse investors who care about all moments of the return distribution can optimally exploit this predictability. Furthermore, we seek to understand if the original findings of predictability are robust and economically significant for such an investor, whether they have vanished in recent years as investors have learned of this predictability and their capabilities of exploiting it have increased.

We answer these questions in the metric of power utility functions with a bootstrapped out-of-sample approach. We use Brandt, Santa-Clara, and Valkanov's (2009) parametric portfolio policy (PPP) to build characteristic-based portfolios that maximize in-sample power utility. We evaluate these portfolios out of sample. We confront estimation risk and model selection with a two-stage dynamically regularized out-of-sample design. We use the first out-of-sample stage to construct

We are grateful to the editor, Clemens Sialm and two anonymous referees for helping us to improve our paper. We retain responsibility for all errors. Send correspondence to Chris Lamoureux, lamoureu@arizona.edu.

The Review of Asset Pricing Studies 14 (2024) 450–480

https://doi.org/10.1093/rapstu/raae006 Advance Access publication 18 March 2024

density functions of all of the optimal portfolios' certainty equivalent returns. We use a max-min procedure to select the optimal portfolio policy. We evaluate this optimal portfolio policy in a second out-of-sample stage. Since both the in- and out-of-sample periods are bootstrapped, we construct (small-sample) empirical distributions and report confidence intervals for functions of interest, such as portfolio alpha, Sharpe ratio, and certainty equivalent return.[1] Evaluating characteristics' predictive efficacy with this loss function, in the expected utility metric, where we consider all moments of portfolios' return distributions, addresses concerns about the statistical robustness and economic relevance of the return predictability. Furthermore, expected return predictability may persist in equilibrium if it is subject to large outliers and/or negative skewness. Barroso and Santa-Clara (2015a) and Kadan and Liu (2014) show that characteristic-based portfolios that generate a high alpha and Sharpe ratio may come at the cost of negative skewness. Nagel (2021, p. 33) notes that "whether methods that deliver the most accurate return forecasts at the individual stock level also automatically give us the best performing portfolio once we aggregate across stocks is an open question that does not have an obvious answer."

We find that over the period 1955–1998, all six of the characteristics that we consider—size, the book-to-market ratio, momentum, average same-month return, residual volatility, and beta—have economically meaningful predictive content for the purpose of forming optimal portfolios from a power utility investor's perspective. Our primary performance metric is the portfolio's certainty equivalent return to a power utility investor with coefficient of relative risk aversion, $\gamma = 2$. In the out-of-sample period, 1990–1998, this investor's regularized dynamically optimized optimal portfolio's certainty equivalent return has a 95% confidence interval of $(329, 529)$ basis points per month compared to the market portfolio's $(121, 137)$.

We analyze overfitting in Brandt, Santa-Clara, and Valkanov's (2009) PPP which has been used successfully in a variety of applications.[2] PPP is parsimonious and avoids the first step in traditional portfolio selection—estimating, or even taking a stand, on the joint distribution of returns conditional on measurable characteristics. Aït-Sahalia and Brandt (2001, p. 1299) characterize this first step as the "Achilles' heel of conditional portfolio choice because although the moments are predictable, this predictability is for some moments quite tenuous." They argue that not specifying a likelihood (i.e., the conditional return distribution) avoids "introducing additional noise and potential misspecifications through the intermediate, but unnecessary, estimation of the return distribution."

[1] Lewellen, Nagel, and Shanken (2010) stress the importance of presenting confidence intervals since sample statistics in asset pricing are often biased and skewed.

[2] DeMiguel, Plyakha, Uppal, and Vilkov (2013) use PPP to examine the predictive content of option-implied moments in a mean-variance setting. Faias and Santa-Clara (2017) analyze optimal option portfolios. Kroencke, Schindler, and Schrimpf (2014) and Barroso and Santa-Clara (2015b) consider foreign exchange portfolio strategies, including the carry trade. Barroso, Reicheneker, Reicheneker, and Rouxelin (2023) optimize jointly over global equity and currency exposure.

Best and Grauer (1991) stress that overfitting causes the documented poor out-of-sample performance of optimal mean-variance portfolios. Optimization amplifies estimation errors. The literature suggests three approaches to mitigate overfitting in portfolio optimization settings. These regularization procedures include: constraining the weights or estimators, Bayesian priors, and machine learning. We use machine learning to manage estimation risk. Jagannathan and Ma (2003) demonstrate that there is a duality between constraining the weights, for example with short-selling constraints, and shrinking moment estimators in the mean-variance optimization context. Bayesian approaches establish a prior using economic theory. Pástor (2000) and Pástor and Stambaugh (2000) use asset pricing models to form the prior. MacKinlay and Pástor (2000) impose moment restrictions according to a factor model. Kan and Zhou (2007) derive the expected loss function from using sample (rather than true) moments when returns are normally distributed. They show that estimation risk can be diversified by holding a minimum variance portfolio in addition to the estimated tangency portfolio. These solutions which effectively reduce portfolio variance relative to the population solution suggest an approach to overfitting more generally by increasing the shadow cost of the return variance (and kurtosis) in terms of mean return (and skewness). DeMiguel, Garlappi, and Uppal (2009) show that in general these attempts to mitigate estimation error in (mean-variance) portfolio selection are dominated by an equally-weighted benchmark (the $\frac{1}{N}$ rule). Barroso, Reicheneker, Reicheneker, and Rouxelin (2023) consider benchmark constraints which limit the deviation of weights from standard equally-weighted and value-weighted benchmark portfolios. Since such constraints tend to reduce portfolio variance they mitigate estimation risk.

Machine learning approaches such as Lasso introduce a hyperparameter, or tuning parameter, to manage estimation risk. For example, DeMiguel, Martín-Utrera, Nogales, and Uppal (2020) use a PPP algorithm to maximize the Sharpe ratio (i.e., they specify a quadratic utility function), with transactions costs. They impose an L1-norm penalty on the parameter space, and demonstrate that it does better out of sample than a non-regularized optimization. Freyberger, Neuhirl, and Weber (2020) use a group Lasso procedure to shrink the model and manage overfitting. Ao, Li, and Zheng (2019) develop a Lasso-type estimator to deal with a large cross-section specifically designed to address the out-of-sample deterioration of the Sharpe ratio. However, Kozak, Nagel, and Santosh (2020, p. 274) note that such a penalty has poor statistical properties when the characteristics are correlated, and it lacks economic motivation. We regularize the PPP by separating the curvature of the loss function that links portfolio weights directly to characteristics from the investor's utility function. Under a power utility function, the coefficient of relative risk aversion, γ is effectively the shadow cost of variance relative to expected returns. We expand the parameter space to allow a power utility investor with coefficient of relative risk aversion γ to increase this shadow cost in sample by maximizing expected utility with coefficient of relative aversion $\gamma^* = \gamma + \lambda$. The hyperparameter $\lambda > 0$ will reduce estimation risk, if present, to the extent that noise is positively linked to the variance of the conditional return-generating process.

We find that for mid-levels of relative risk aversion PPP does not suffer from estimation risk, lending credence to the claim that estimating moments of the conditional return distribution is the source of much overfitting (Aït-Sahalia and Brandt 2001). However, estimation risk is a serious problem for PPP to our power utility investor with a coefficient of relative risk aversion of two. Since the PPP is agnostic with respect to the conditional return-generating process, we cannot appeal to Bayes' Theorem to manage estimation risk. Instead, we rely on the multiprior decision theory of Gilboa and Schmeidler (1989). Gilboa and Schmeidler (1989, p. 142) consider uncertainty—as distinct from risk—where "there is too little information to form a prior." They show that uncertainty aversion means that the agent should optimize over all feasible states and choose that rule which produces the best outcome under the worst possible state of nature.

Because there is no likelihood we use the bootstrap to construct the sampling distribution of out-of-sample portfolio properties for each model configuration. A configuration consists of the curvature of the loss function used to estimate an optimal portfolio rule in-sample (λ) and the (sub)set of measurable characteristics. With six characteristics there are $63 = \sum_{j=1}^{6} \frac{6!}{(6-j)! \cdot j!}$ unique combinations. We consider 14 values of λ. So we evaluate all $63 \times 14 = 882$ alternative configurations at the beginning of each year in our out-of-sample periods. After (minimally) 15 years of out-of-sample data, we evaluate the utility function of these out-of-sample returns. We now have a bootstrapped sampling distribution of the utility function of the out-of-sample returns from each configuration. Confronted with a finite sample, the investor seeks to maximize expected utility in the worst-case scenario (i.e., max-min). We select that configuration with the highest 1 percentile value of the loss function (certainty equivalent return) at the beginning of each year in the second-stage out-of-sample period to construct the bootstrap distribution of the returns on the dynamically optimal portfolio policy.[3] This is linked to statistical assessment of the portfolio. We consider that Portfolio A dominates (i.e., is statistically significantly strictly preferred to) Portfolio B if: a) A's 2.5 percentile certainty equivalent return is greater than B's 97.5 percentile certainty equivalent return; and/or b) A's 2.5 percentile certainty equivalent return is positive but B's 2.5 percentile certainty equivalent return is negative.

Our use of the max-min criterion on the bootstrapped out-of-sample certainty equivalent returns relates to the literature in decision making under uncertainty with machine learning. We learn from specifications' out-of-sample performance, as in Barroso and Saxena (2022) and Freyberger, Neuhierl, and Weber (2020). Gilboa, Postlewaite, and Schmeidler (2008) provide an overview and survey of the problem of decision making under uncertainty, and the "multiple prior"

[3] We use the bootstrap 1 percentile as the "worst-case scenario" to accommodate numerical issues and link to statistics. Our results do not change in any qualitative way if we use the 2.5 percentile instead of the 1 percentile certainty equivalent to select the optimal out-of-sample configuration.

approach. Aït-Sahalia and Brandt (2001) suggest the use of max-min for a Constant Relative Risk Aversion (CRRA) investor in the case where the (conditional) return distribution is unknown. This approach has been extended broadly to dynamic optimization by Hansen and Sargent (2008). There is related work in operations research and machine learning on robust optimization, where "it is assumed that the decision maker has no distributional knowledge about the underlying uncertainty except for its support, and the model minimizes the worst-case cost over an uncertainty set," (Rahimian and Mehrotra 2022, p. 1). Bertsimas, Gupta, and Kallus (2018) characterize this approach as "data-driven robust optimization."

Nagel (2021, p. 48) notes that, "shrinkage can improve portfolio performance if there is heterogeneity in the covariates' relative contribution to moments and estimation error. Shrinkage must reduce undesirable contributions (estimation error, variance, and kurtosis) more than desirable ones (return mean and skewness)." We regularize or shrink estimation by disentangling the loss function maximized on the data to obtain portfolio weights from the investor's utility function. Our two findings, that $\lambda > 0$ greatly reduces estimation risk for our primary (relatively risk-tolerant) investor and $\lambda = 0$ for more risk-averse investors, suggest that estimation risk, the tendency to find spurious patterns in a sample, is related to the variance of the portfolio return distribution. This is fully consistent with all of the literature that demonstrates that constraining portfolio leverage (hence variance) serves to mitigate estimation risk.

We find that characteristics' predictive usefulness for portfolio selection is temporally unstable, and has vanished post-1998. A jump in λ under the rolling protocol prior to 2001 accompanies this structural break. Our max-min regularization suggests more conservative portfolios following out-of-sample results that are disappointing compared to prior expectations. This result is consistent with recent research. For example, Martin and Nagel (2022) provide a learning framework in which an econometrician can detect such cross-sectional predictability using standard tests, but the predictability is no longer present out of sample. They motivate the use of an out-of-sample test design–noting that we should expect to find evidence of predictability in sample in a high dimensional highly complex environment. In this setting, the usual implications of informationally efficient markets place testable restrictions on out-of-sample predictability. Green, Hand, and Zhang (2017) document a significant drop in characteristics' predictive content in 2003, which they attribute to institutional changes that reduce trading frictions.[4] McLean and Pontiff (2016) also document a drop in the return to trading on anomalies documented in the literature subsequent to publication. Our result, that measurable characteristics did have economically and statistically predictive content for portfolio construction prior to 1999 but no longer do, complements this research.

[4] These changes stem from both regulations: Regulation FD (2000) and Sarbanes-Oxley (2002); and technological advances: decimalization (2001) and enhanced autoquote (2003).

We consider power utility investors with higher aversions to risk, γ values of 5 and 8. The corresponding tables and figures are provided in an Internet Appendix. We confirm the findings that CRRA investors could use PPP to exploit the predictive content in characteristics prior to 1999, but not since then. As noted, while overfitting is a severe problem for the more risk-tolerant ($\gamma = 2$) investor in this first subperiod it is not for the more risk-averse investors. In the first subperiod optimal portfolio variance declines statistically significantly in risk aversion. These results are consistent with the hypothesis that estimation risk shrinks in portfolio variance. However, optimal portfolio Sharpe ratio, skewness, and kurtosis are flat in risk aversion.

We complement studies by Lewellen (2015), Green, Hand, and Zhang (2017), and Freyberger, Neuhirl, and Weber (2020) by showing how the set of characteristics affects optimal portfolios' factor exposures in this first subperiod. Our most risk-tolerant investor uses the characteristics to short the market, and get positive exposures to the Fama-French value, size, and momentum factors. Roughly half of the portfolio's excess mean return and return variance come from outside the span of the six Fama-French factors. The portfolio has a small positive exposure to their Robust Minus Week (RMW) factor, and a statistically significant negative loading on their Conservative Minus Aggressive (CMA) factor. While the size of characteristic weight tilts diminish in risk aversion, the percentages of portfolio mean returns and return variance within the span of the Fama-French six-factor model are stable in γ.

1. Portfolio Selection

1.1 Algorithm

In Brandt, Santa-Clara, and Valkanov's (2009) algorithm, the vector θ is estimated to maximize a concave loss function over M periods:

$$\max_{\theta} \sum_{m=0}^{M-1} \frac{(1 + r_{p,m+1})^{1-\gamma^*}}{1-\gamma^*} \left(\frac{1}{M}\right), \tag{1}$$

by allowing portfolio weights to depend on observable stock characteristics:

$$r_{p,m+1} = \sum_{i=1}^{N_m} \left(\bar{\omega}_{i,m} + \frac{1}{N_m} \theta' x_{i,m}\right) \cdot r_{i,m+1}, \tag{2}$$

where: $x_{i,m}$ is the K-vector of cross-sectionally standardized characteristics on firm i, measurable at month m; $\bar{\omega}_{i,m}$ is the weight of stock i in the (value-weighted) market portfolio at month m; and N_m is the number of stocks in the sample in month m. Conditioning only on information that is available to investors at the time the portfolios are formed avoids the overconditioning bias analyzed by Boguth, Carlson, Fisher, and Simutin (2011). Unlike Brandt, Santa-Clara, and Valkanov (2009) or DeMiguel, Martín-Utrera, Nogales, and Uppal (2020), we do not identify γ^*, the parameter used to generate a feasible portfolio strategy in

(1), with the relevant statistical loss function (i.e., a specific investor's utility function). Instead, we consider a statistical loss function (alternatively "an investor") that takes the same form as (1) indexed by γ (the curvature of the statistical loss function which is pre-determined and fixed). From this perspective, γ^* is a choice variable—a tool to manage estimation risk. Letting $\gamma^* \equiv \gamma + \lambda$, λ is a hyperparameter, or tuning parameter, of shrinkage or regularization.[5]

Our primary focus is on a power utility investor with coefficient of relative risk aversion, $\gamma = 2$. We also consider the effects of increasing risk aversion on estimation risk and the predictive content of characteristics for portfolio formation. Our interest is in out-of-sample statistical comparisons across portfolios generated by various feasible portfolio rules—from the perspective of this nonlinear statistical loss function. An eligible portfolio rule is a configuration consisting of a subset of characteristics and the increased penalty on variance and kurtosis, expressed in utility terms, λ. By definition, the optimal in-sample portfolio is obtained using all characteristics and by setting $\lambda = 0$, (i.e., $\gamma^* = \gamma$). Whether this is also true out of sample is an empirical question as it depends on the unmodeled relationship between estimation error, the characteristics, and the loss function. If there is a positive relationship between an optimal portfolio's in-sample variance and its noise (i.e., higher out-of-sample variance), then using a more concave loss function than the investor's utility function ($\lambda > 0$) may generate portfolios that are preferred to those obtained by constraining γ^* to equal γ.

1.2 Data

Because our model selection stage uses out-of-sample analysis of expected utility from hundreds of configurations we require a comparatively long time series. Therefore, we use characteristics that can be directly computed from market prices, and the book value from the CRSP-Compustat merged file. Our sample of returns and characteristics spans the years $Y_1 = 1960$ through $Y_{62} = 2021$, which means we use complete data starting in January 1955 to obtain all measurable characteristics as of the start of 1960. The initial in-period estimation uses the 180 months January 1960 - December 1974, and our initial out-of-sample validation period is 1975–1989. Our fully out-of-sample period comprises the 32 years 1990–2021. We use Y_y to indicate year Y, for $y = 1, \ldots, 62$, and $m = 1, \ldots, M = 744$ to denote the month in our sample.

For a stock to be eligible for investment in month m, we require five years of (non-missing) returns in months $[m - 60,\ m - 1]$. If the stock return is missing in month m, we look to the CRSP delisting return. If that is missing, we substitute -30% for stocks listed on the New York and American stock exchanges and -50% for Nasdaq stocks. Thus, the stocks in the January 1960 sample have no

[5] This algorithm is general and can be used with many alternative specifications. One extension that Brandt, Santa-Clara, and Valkanov (2009, Section 1.3.3) mention is allowing weights to be nonlinear functions of characteristics. Freyberger, Neuhirl, and Weber's (2020) finding that non-linear functions of characteristics improve the Sharpe ratio relative to a linear specification rationale for such an extension.

missing data from January 1955 through December 1959. These requirements limit the sample. For example, Brandt, Santa-Clara, and Valkanov (2009) report that the smallest cross-section in their study comprises 1,033 stocks in February 1964. Only 624 firms satisfy our data requirements in that month. Prior research suggests that it is important to exclude penny stocks and stocks with low relative market capitalization. To this end we add two additional criteria for a stock to be eligible for inclusion in month m. First, to exclude nano and small microcap stocks we impose a real dollar minimum equity market capitalization in month $m-1$ of \$110 million in December 2021 dollars. We obtain the US Consumer Price Index from the Federal Reserve (FRED).[6] This restriction excludes stocks with market capitalization less than \$11.5 million in January 1960 and \$50 million in January 1990. Second, we exclude the smallest 10% of qualifying stocks in the months before Nasdaq stocks enter our sample, which is January 1978. We exclude 20% of eligible firms when Nasdaq stocks enter the sample. In February 1964, the dollar criterion discards 39 of the eligible 624 stocks, and the percent exclusion discards another 58 stocks. This leaves a final sample of 527 stocks—51% of Brandt, Santa-Clara, and Valkanov's cross-section on that date.

Table IA-1 in the Internet Appendix provides details on the sample. For each of the 744 months in the full sample, the table shows: prior-month end, the number of stocks that meet the data availability requirement for month m (at month $m-1$), the number excluded by the minimum equity market capitalization constraint, and the final sample size, along with the minimum and median market capitalizations in the sample. There are 411 (exclusively New York Stock Exchange) stocks in the final sample in January 1960. The sample size jumps in August, 1967, from 678 to 881 when stocks listed on the American Stock Exchange are eligible for inclusion in our sample. The largest jump in sample size is in January 1978 (from 1,001 to 1,420 stocks) when Nasdaq stocks are eligible to enter our sample. The maximum number of stocks is 2,290 in April, 2006. There are 1,848 stocks in our sample in August 2008 and 1,693 stocks in the last month in our sample, December 2021.

We use the following six characteristics: momentum (ζ), the book-to-market ratio (V), log size (S), beta (β), market model residual standard deviation (σ_ϵ), and the average same-month return over the preceding five years ($\bar{r}$).[7] Momentum is the stock's compounded return over the annual period $[m-13,\ m-2]$. Equity market capitalization is the market value of the company's outstanding shares (aggregated across all share classes) at time $m-2$. Book value is obtained from the Compustat database for the most recent fiscal year-end between $m-6$ and $m-18$. We define the book-to-market ratio to be the natural log of one plus the ratio of book value to equity market capitalization. Size is equity market capitalization. We estimate beta and the residual standard deviation by regressing monthly returns over the period $[m-60,\ m-1]$ on the CRSP value-weighted index.

6 We extracted the series CPIAUCSL: consumer price index for all urban consumers: all items.

7 Same-month seasonality is analyzed in Heston and Sadka (2008).

We normalize and standardize the characteristics so that each characteristic has a cross-sectional mean of zero and unit standard deviation. Inspection of (2) shows that optimal portfolio weights will thereby sum to unity for any value of θ. This also means that the characteristics are observationally equivalent to shrinkage values. For example, let β be a stock's OLS beta. Consider a prior-weighted beta, such as $\beta^S = .5 \cdot \beta + .5 \cdot 1$. The normalized β^S are identical to the normalized β. This implicit shrinkage mitigates the usual concerns about outliers in characteristic space so we do not winsorize the characteristics. A single observation ($\Psi_{i,m}$) comprises stock i's return in month m, $r_{i,m}$, as well as the vector of characteristics, measurable at month $m-1$, for stock i, $i = 1, \ldots, N_m$. Importantly, $\Psi_{i,m}$ also includes stock i's market capitalization at $m-1$, since the passive portfolio at m (i.e., the portfolio when the θ-vector is zero) is the market-weighted portfolio at $m-1$.

We are interested in the optimal sets of characteristics so we consider all possible sets using these six characteristics. There are 63 such sets including each characteristic as a singleton and all six as a sextuplet.

1.3 Bootstrap sample construction

In light of the high noise-to-signal nature of stock-return data, we develop and implement a bootstrap design for optimization, regularization, and to characterize the out-of-sample sampling distributions of portfolio properties, such as certainty equivalent, portfolio loading on factors, portfolio skewness, etc. We stack the 744 row-vectors $\Psi_{i,m}$ (of varying lengths) to form the array $\mathbf{\Psi}$. This is not a traditional panel since stock i at month m is different from stock i at any other time, and as noted, the number of columns is different for each row.

Our resampling is nonparametric cross-sectional bootstrap (see Kapetanios 2008), motivated by the (repeated) single-period optimization problem with out-of-sample cross-validation, regularization, and inference. It is nonparametric since we draw from the raw data. It assumes that returns are temporally independent conditional on $\mathbf{\Psi}$. As such, $\mathbf{\Psi}^{\boldsymbol{b}}$, is the b^{th} resample from $\mathbf{\Psi}$, for $b = 1, \ldots, B =$10,000. Consistency holds under $N-$asymptotics (as the number of stocks in each period increases), since we can view $\widehat{\theta}$ as a GMM estimator. Preserving the original time series in all resamples allows us to evaluate the effects of structural instability on out-of-sample portfolios, and also maintains the cyclical (month of the year) and serial dependence patterns inherent in the design of $\mathbf{\Psi}_{\boldsymbol{m}}$, for $m = 1, \ldots, M = 744$. A resampled draw for month m resamples (with replacement) an N_m-row vector from $\mathbf{\Psi}_{\boldsymbol{m}}$. Thus, each pseudosample consists of the same number of observations in each period as the original sample and we preserve the calendar structure of the original data. We also maintain the temporal structure of the investment opportunity set. In the original data $\mathbf{\Psi}_{\boldsymbol{m}}$ includes the raw values of the characteristics as well as the market weight of the stock at $m-1$.

Once we draw a resampled cross-section from $\mathbf{\Psi}_{\boldsymbol{m}}$, we normalize (so that all characteristics have a zero mean in this month in this bootstrapped sample), standardize (so that all characteristics have unit variances), and construct the value-weighted market weights for each resampled stock, based on the total market capitalization in this month in this bootstrapped sample. Thus, our resampling design preserves the integrity of the investment opportunity set at each *m*. This critically includes the relationship between stocks' characteristics and sizes. We also maintain the integrity of the seasonality in the data. The month of January, for example, occurs exactly once in any 12-month cycle, and its timing is deterministic. The matrix of characteristics at *m* is not independent of the size of each stock at *m*. We manage concerns about unmodeled temporal dependencies by maintaining the out-of-sample test design with each bootstrapped sample.

1.4 Dynamic regularization

We implement a dynamic two-stage bootstrapped out-of-sample regularization process using both updating and rolling construction. We provide an appendix containing pseudo code for our approach. Our first stage entails construction of the bootstrap distribution of out-of-sample returns from each of the 882 configurations. A configuration is a characteristic set and a value for λ (or γ^*). As noted above, with six characteristics, and 14 values of $\lambda = \{0, 1, \ldots, 11, 14, 20\}$, we consider 882 configurations under each protocol. In each of the $b = 1, \ldots, 10,$ 000 bootstrap samples $\boldsymbol{\Psi}^{\boldsymbol{b}}$, we estimate the parameter vector $\widehat{\theta}$ *in sample* by maximizing (1) over the K-vector θ for the in-sample period using a modified Newton method.[8] The first in-sample period (under both protocols) is 1960–1974. We use the $\widehat{\theta}$ estimated from this period to form the out-of-sample portfolio in each of the next 12 months (in 1975). Under the updating protocol, then the 12 months of 1975 are added to the sample, and $\widehat{\theta}$ is estimated on this 192-month in-sample period. This $\widehat{\theta}$ is used to construct out-of-sample portfolio returns in the 12 months of 1976. Under the rolling protocol we drop the first 12 months from the in-sample data so that the second year's θ is estimated using the 180-month in-sample period 1961–1975. At this point we have 10,000 bootstrap draws of 24 months of out-of-sample returns from each of the 882 configurations.

Our dynamic regularization selects the optimal configuration by looking back at the performance of prior out-of-sample returns. We use at least 15 years of optimal portfolio out-of-sample returns for our second-stage regularization, so the first year of second-stage (i.e., dynamically regularized) out-of-sample returns is 1990. At the end of 1989 we have 180 months of 10,000 draws of out-of-sample optimal portfolio returns from each of the 882 configurations. We select that configuration whose bootstrap distribution of out-of-sample returns over the period 1975–1989 has the highest 1 percentile certainty equivalent return as ex post optimal, and use this to construct the dynamically optimized (and regularized) out-of-sample returns for the next 12 months (1990). From this point until the end of the sample, each additional year involves re-estimating the parameter vector using the

[8] We use the model/trust region algorithm of Gay (1983), and the analytical gradient and Hessian from (1). Our two-stage algorithm is numerically intensive. We maximize in-sample utility 414,540,000 times.

Table 1

Optimal γ^* and portfolio (characteristic sets): Sampling properties of certainty equivalent returns for investor with power utility and coefficient of relative risk aversion, $\gamma = 2$

	Updating protocol					Rolling protocol				
Next	Optimal		Certainty Equivalent			Optimal		Certainty Equivalent		
Year	Chars	γ^*	1%ile	Mean	Std Dev	Chars	γ^*	1%ile	Mean	Std Dev
1990	**VWI**		113.3	120.7	3.2	**VWI**		113.3	120.7	3.2
1990	**EWI**		145.0	149.1	1.8	**EWI**		145.0	149.1	1.8
1990	ζ,V,S,$\bar{r}$,σ_ϵ	2	−10,000.0	−4,146.6	5,245.0	ζ,V,S,$\bar{r}$,σ_ϵ	2	−10,000.0	−437.6	3,148.4
1990	ζ,V,S,$\bar{r}$,σ_ϵ	3	527.8	625.5	46.1	ζ,V,S,$\bar{r}$,σ_ϵ	3	523.1	622.2	46.0
1991	ζ,S,$\bar{r}$,σ_ϵ	3	497.7	589.7	43.5	ζ,S,β,$\bar{r}$,σ_ϵ	3	498.4	597.8	46.1
1992	ζ,S,$\bar{r}$,σ_ϵ	3	512.1	603.4	43.4	ζ,S,β,$\bar{r}$,σ_ϵ	3	497.4	601.6	48.1
1993	ζ,V,S,β,$\bar{r}$,σ_ϵ	3	512.3	615.5	48.1	ζ,V,S,β,$\bar{r}$,σ_ϵ	3	486.8	597.4	51.3
1994	ζ,V,S,β,$\bar{r}$,σ_ϵ	3	518.2	619.6	47.3	ζ,V,S,β,$\bar{r}$,σ_ϵ	3	493.3	606.4	52.3
1995	ζ,V,S,β,$\bar{r}$,σ_ϵ	3	500.8	601.9	46.2	ζ,V,S,β,$\bar{r}$,σ_ϵ	3	471.1	593.3	54.2
1996	ζ,V,S,β,$\bar{r}$,σ_ϵ	3	489.7	584.7	44.2	ζ,V,S,β,$\bar{r}$,σ_ϵ	3	496.1	615.5	54.3
1997	ζ,V,S,β,$\bar{r}$,σ_ϵ	3	477.8	568.1	42.6	ζ,V,S,β,$\bar{r}$,σ_ϵ	3	503.1	620.6	54.5
1998	ζ,V,S,β,$\bar{r}$,σ_ϵ	3	503.9	598.6	44.3	ζ,V,S,β,$\bar{r}$,σ_ϵ	3	540.8	670.4	60.4
1999	ζ,V,S,β,$\bar{r}$,σ_ϵ	3	477.0	565.3	42.2	ζ,V,β,$\bar{r}$,σ_ϵ	3	488.7	624.1	61.9
2000	ζ,V,S,β,$\bar{r}$,σ_ϵ	3	472.2	562.5	42.9	ζ,V,β,$\bar{r}$,σ_ϵ	3	485.7	643.8	240.0
2001	ζ,S,β,$\bar{r}$,σ_ϵ	3	443.8	530.1	45.3	ζ,S,β,$\bar{r}$,σ_ϵ	7	229.4	322.9	39.9
2002	ζ,S,β,$\bar{r}$,σ_ϵ	3	415.5	503.0	44.4	ζ,S,β,$\bar{r}$,σ_ϵ	8	199.7	278.5	35.1
2003	ζ,S,β,$\bar{r}$,σ_ϵ	3	432.4	521.3	44.9	ζ,S,β,$\bar{r}$,σ_ϵ	8	219.1	303.6	38.1
2004	ζ,S,β,$\bar{r}$,σ_ϵ	3	401.5	483.8	41.7	ζ,S,β,$\bar{r}$,σ_ϵ	8	200.9	283.1	36.6
2005	ζ,S,β,$\bar{r}$,σ_ϵ	3	395.8	476.4	40.6	ζ,S,β,$\bar{r}$,σ_ϵ	8	181.7	265.6	36.8
2006	ζ,S,β,$\bar{r}$,σ_ϵ	3	388.9	468.3	39.5	ζ,S,β,$\bar{r}$,σ_ϵ	8	177.9	260.1	36.3
2007	ζ,S,β,$\bar{r}$,σ_ϵ	3	374.5	450.9	37.9	ζ,S,β,$\bar{r}$,σ_ϵ	8	154.8	235.2	35.0
2008	ζ,S,β,$\bar{r}$,σ_ϵ	3	345.9	415.7	35.4	ζ,S,β,$\bar{r}$,σ_ϵ	9	115.0	181.5	28.2
2009	ζ,S,β,$\bar{r}$,σ_ϵ	3	339.8	410.5	35.1	ζ,S,β,$\bar{r}$,σ_ϵ	10	87.7	146.7	26.0
2010	ζ,S,β,$\bar{r}$,σ_ϵ	3	305.7	376.1	35.1	**EWI**		72.8	77.6	2.1
2011	ζ,S,β,$\bar{r}$,σ_ϵ	3	306.6	376.0	34.4	**EWI**		69.2	74.1	2.1
2012	ζ,S,β,$\bar{r}$,σ_ϵ	3	306.2	375.4	34.1	ζ,S,β,$\bar{r}$,σ_ϵ	9	58.6	128.2	30.0
2013	ζ,S,β,$\bar{r}$,σ_ϵ	3	299.6	366.8	33.0	**EWI**		50.3	55.4	2.1
2014	ζ,S,β,$\bar{r}$,σ_ϵ	3	301.4	367.4	32.4	**EWI**		69.7	74.6	2.1
2015	ζ,S,β,$\bar{r}$,σ_ϵ	3	284.5	348.2	31.0	**EWI**		65.9	70.5	2.0
2016	ζ,S,β,$\bar{r}$,σ_ϵ	3	279.0	339.6	29.9	S,$\bar{r}$,σ_ϵ	6	84.8	114.0	13.8
2017	ζ,S,β,$\bar{r}$,σ_ϵ	3	273.7	333.0	29.1	S,$\bar{r}$,σ_ϵ	7	77.2	96.2	9.1
2018	ζ,S,β,$\bar{r}$,σ_ϵ	3	260.7	317.3	27.8	**EWI**		81.0	85.1	1.8
2019	ζ,S,β,$\bar{r}$,σ_ϵ	4	249.3	295.1	20.9	S,σ_ϵ	16	63.0	73.3	4.5
2020	ζ,S,β,$\bar{r}$,σ_ϵ	4	243.1	287.3	20.3	S,σ_ϵ	22	68.7	80.0	4.9
2021	ζ,S,β,$\bar{r}$,σ_ϵ	4	236.3	278.6	19.3	V,S,σ_ϵ	22	69.5	84.5	6.5

Weight tilts (θ) are estimated for 63 characteristic sets under each of 14 values of the loss function curvature (γ^*) using both rolling and updating protocols. Of these 882 cases that with the highest 1 percentile value of the out-of-sample certainty equivalent is reported for each of three investors in basis points per month. The characteristic symbols are: ζ, momentum; V, book-to-market ratio; S, log size; β, from lagged 60-month market model; $\bar{r}$, average same-month return over the previous five years; σ_ϵ, standard deviation of lagged 60-month market model residual.

additional year's worth of data (as in Stage 1 optimization). Further, we (dynamically) regularize using the additional year of out-of-sample returns. We proceed in this manner through the end of 2020 to construct the dynamically optimized (and regularized) out-of-sample returns for 2021. The optimal configurations at the end of each year from both protocols are reported in Table 1 (and in Tables IA-2 and IA-3 for the dynamically optimized portfolios for the increasingly more risk-averse power utility investors). At any month in the second out-of-sample stage, the optimal configuration and its corresponding $\widehat{\theta}$ vector are determined using information at the end of the previous year. The stocks' characteristics (and market weights) are available prior to the start of the month.

As in Brandt, Santa-Clara, and Valkanov (2009), we estimate the $\widehat{\theta}$ coefficients at the beginning of each year in the out-of-sample period. There are several reasons for this. Since the updating protocol relies on the temporal structural stability of the return-generating process, adding 12 months of data increases the reliability and efficiency of the θ estimates. By contrast, since the rolling protocol uses only the most recent 15 years of data, if the conditional return-generating process experiences structural breaks, that will be evident in material changes in $\widehat{\theta}$ over time. We similarly choose the optimal configuration using the max-min criterion at the beginning of each year in the second-stage out-of-sample period. The rolling protocol accommodates changes in the joint return estimation error process. This dynamic regularization is consonant with the annual updating of $\widehat{\theta}$ and is consistent with an investor updating her information set as she moves through time. In this design, that new information is used in two ways: first, to update the $\widehat{\theta}$ vector, and second, to select the optimal configuration. As the Appendix shows, the sampling protocols differ in both the in-sample period used to obtain $\widehat{\theta}$ and the period used in second-stage model selection.

2. Results

Table 1 shows the ex post optimal portfolio policy for the power utility investor with $\gamma = 2$ at the beginning of each year in the second-stage out-of-sample period under both protocols. This is the optimal configuration from the first-out-of-sample period that ends before the indicated year: the characteristic set (of the 63 possibilities) and $\gamma^* = \gamma + \lambda$ (14 possibilities). The table provides sampling statistics (1 percentile value, mean, and standard deviation) of this configuration's certainty equivalent from this period. Table 1 provides the same information about the sampling distribution of the two benchmarks' certainty equivalents at the beginning of the second-stage out-of-sample period. The benchmarks are the value-weighted portfolio of all eligible securities at the beginning of each month (VWI) and the equally-weighted portfolios of all eligible securities at the beginning of each month (EWI). The table also reports the certainty equivalent sampling distribution for the configuration with the optimal characteristic set and $\lambda = 0$ at the end of the first out-of-sample period under both protocols. Figure 1 reports the sampling distributions (box and whiskers plots) of the θ coefficient on each of the six characteristics for each year in the second-stage out-of-sample period from the rolling protocol. The whiskers show the 95% confidence interval on the estimate each year, and the box the interquartile range. The median is the bar inside the box.

Table 1 shows that $\lambda > 0$ is a necessary regularization for in-sample optimization to produce attractive out-of-sample returns for this investor under both protocols. This suggests that estimation risk is positively related to the variance of the conditional return distribution as $\lambda > 0$ penalizes variance (and kurtosis) more

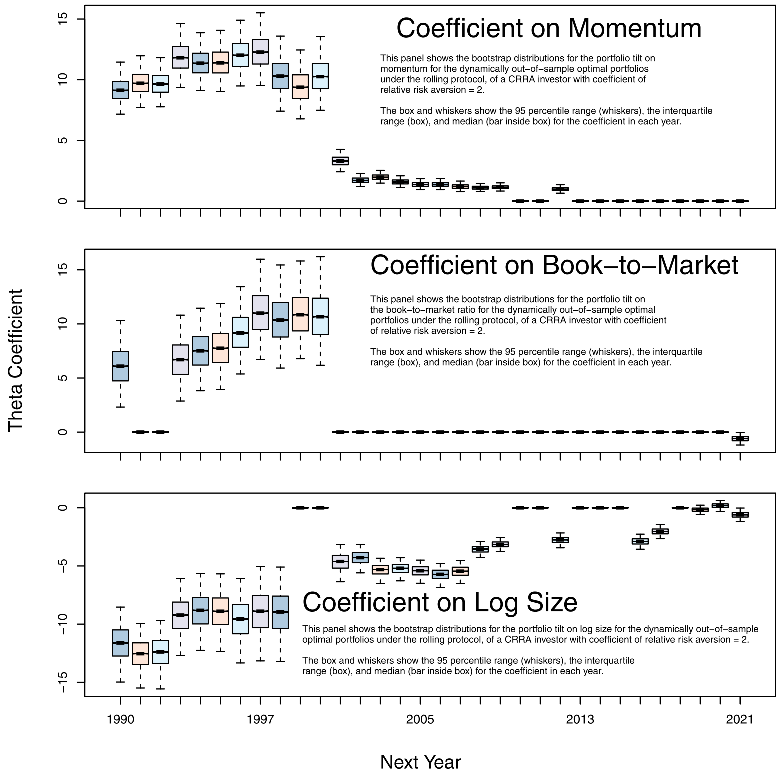

Figure 1
Characteristic-tilts (θ) over time
Sampling distributions of the θ coefficient on (standardized) characteristics from the optimal model over the preceding 180 out-of-sample months, used to construct the optimal portfolio in the indicated year.

than the investor's utility function. This form of regularization has economic rationale—unlike penalizing the θ coefficients. It is also more statistically appealing since the characteristics' effects on portfolio returns are not independent, as noted by Kozak, Nagel, and Santosh (2020).[9]

Table 1 shows that this power utility investor's mean (95% confidence interval) out-of-sample certainty equivalent return without regularization (i.e., with $\lambda = 0$), over the 180 months ending in 1989 is -41% (-100%, 52%) per month under the updating protocol and -4% (-100%, 31%) per month under the rolling protocol. The power utility function is indeterminate at returns less than or equal to -100%. Therefore, we define the certainty equivalent return for a pseudosample to be -100% ($-10{,}000$ basis points) per month if the return in any month in the

9 For example, the θ coefficient on a characteristic may be centered close to 0 but takes on large positive values when the coefficients on other characteristics are small and takes on sizeable negative values when the coefficients on those other characteristics are large. This is especially important in light of the large sampling variation on all characteristics' θ estimates.

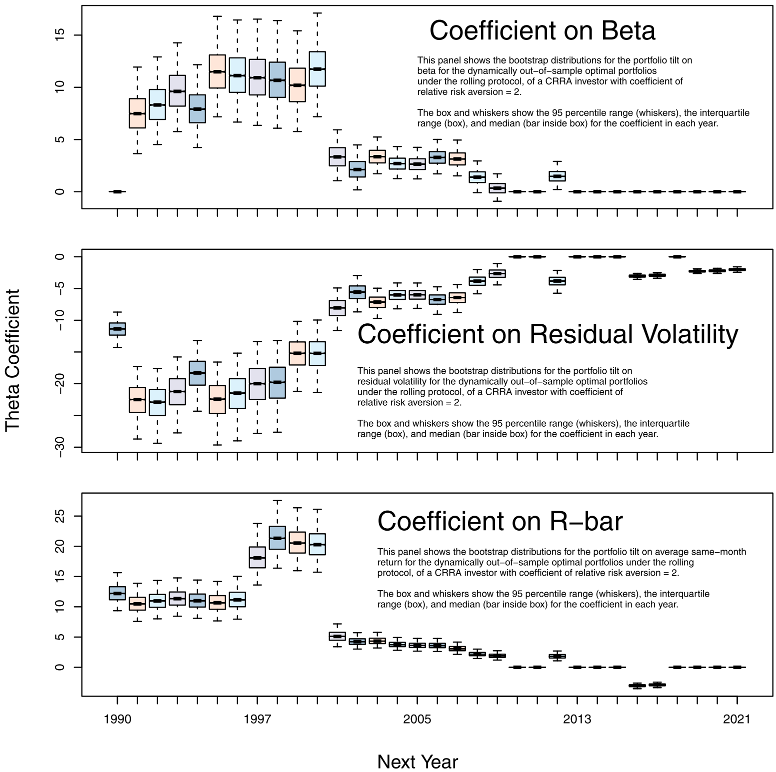

Figure 1
Continued

relevant period does not exceed -1. The mean (95% confidence interval) of the θ coefficients in this case are: momentum, 5.8 (-30.5, 10.8); book-to-market ratio, 5.3 (-2.7, 9.5); log size, -16.5 (-37.5, 10.9); residual volatility, -7.1 (-13.8, 41.1); and average same-month return, 10.5 (-30.5, 10.8). The enormous sampling variation in the weight coefficients shows the reason for the poor performance of this unregularized estimation. The asymmetry in the sampling distributions results from the correlatedness amongst the characteristics and cross-effects. This sampling variation in itself is enough to warn off investors from this configuration.

This type of estimation risk arises because of large absolute sampling correlations between the θ coefficients. In this case, the sampling correlations between the θ coefficient on residual volatility and the θ coefficients on momentum, average same-month return, and log size are -93%, -88%, and -86%, respectively. The sampling correlations between the θ coefficient on momentum and the θ coefficients on average same-month return and log size are 88% and 86%, respectively. As such, this type of overfitting cannot be solved by restricting some of the θ coefficients to zero (e.g., using an L1-norm penalty on θ). None of the 63 characteristic sets generates a portfolio that dominates the benchmark with $\lambda = 0$. By increasing the implicit penalty on the conditional return variance, we are able to take advantage of the cross-effects amongst the characteristics.

Since the utility (statistical loss) function depends on higher moments, we report sampling distributions of measures of skewness and kurtosis. Our robust measures of skewness (S_4) and kurtosis (K_3) are recommended by Kim and White (2004):

$$S_4 = \frac{\mu - r_{.5}}{\sigma} \tag{3}$$

and

$$K_3 = \left\{ \frac{\bar{r}^+_{.95} - \bar{r}^-_{.05}}{\bar{r}^+_{.5} - \bar{r}^-_{.5}} - 2.63 \right\} \cdot 100, \tag{4}$$

where: μ is the mean portfolio return over the out-of-sample period, σ is the return standard deviation, $\bar{r}^+_{.95}$ is the mean of the highest 5% of returns, $\bar{r}^-_{.05}$ is the mean of the smallest 5% of returns, $\bar{r}^+_{.5}$ is the mean of the top half of returns, $\bar{r}^-_{.5}$ is the mean of the bottom half of returns, and $r_{.5}$ is the median out-of-sample return.[10]

2.1 Out-of-sample regularization: Ex post optimal configurations

The optimal configurations are very similar between the two protocols through the year 2000. Both specify $\lambda = 1$ and for the most part use all six characteristics. This similarity is somewhat surprising in light of the large sampling variation in the optimal portfolios' certainty equivalent returns. Many of the 882 alternative configurations have similar certainty equivalent distributions. For example, the optimal configuration under the updating portfolio for 1992 uses only the four characteristics: momentum, log size, average same-month return, and residual volatility, along with $\lambda = 1$. This configuration's 1 percentile out-of-sample certainty equivalent return over the preceding 204 months is 512 basis points per month. The 1 percentile of the configuration with $\lambda = 1$ and all six characteristics (i.e., add the book-to-market ratio and beta to the optimal set) is 509 basis points per month. The configuration with $\lambda = 2$ and all six characteristics has an analogous 1 percentile of 452 basis points per month. Both characteristic sets with $\lambda = 0$ have 1 percentile out-of-sample certainty equivalents of -100% per month.

Figure 1 shows the sampling distributions of the θ coefficients from the ex post optimal portfolios under the rolling protocol preceding each year in the out-of-sample period. The figure illustrates the apparent structural change in the conditional return-generating process around the year 2000. Prior to that point, the θ coefficients on all six characteristics tend to be large in absolute value, and their 95% confidence intervals are far from zero as well.

[10] As per Kim and White (2004), in our pseudosamples S_4, the Pearson skewness coefficient is similar to, and more reliable than S_3, the Bowley skewness coefficient-integrated over the tail size. Similarly K_3, the Hogg coefficient, is more reliable than K_4, the Crow and Siddiqui parameter.

These optimal portfolios tend to tilt the portfolio most aggressively to low residual volatility stocks. They also tilt toward high-beta stocks.[11] Prior to 2000, these ex post optimal portfolios also tilt toward stocks whose average same-month return is high (in the next month) and away from those with relatively low average same-month return over the previous five years. The ex post optimal portfolios also tilt toward stocks that have relatively high returns over months [-13, -2], small stocks, and value stocks (those with relatively high book-to-market values).

The rolling protocol results in Table 1 show that the PPP's performance starts to deteriorate around 1999. Around this time the variance of the ex post optimal portfolio's certainty equivalent return, with $\lambda = 1$, increases four-fold, and then λ increases to 5 and 6. Furthermore, the table shows that the optimal portfolio for the years 2010, 2011, 2013, 2014, 2015, and 2018, is the equally-weighted index of all sample stocks. This means that the 1 percentile value of the certainty equivalent return over the preceding 15 years on all 882 configurations is less than this index's 1 percentile certainty equivalent in that period, which is reported in the table. There is no evident breakdown under the updating protocol, as its ex post optimal configurations are remarkably stable by comparison.

2.2 Optimal regularized portfolios' out-of-sample performance

2.2.1 Subperiod 1 (1990–1998). Table 2 shows that both regularized updating and rolling protocols generate out-of-sample certainty equivalent returns that are statistically and economically significantly larger than the benchmarks in the first subperiod. The updating protocol produces a mean (95% interval) certainty equivalent of 428 $(329, 529)$ basis points per month compared to the value-weighted index of the stocks in the universe of 129 $(121, 137)$. The optimal portfolios' Sharpe ratio means (95% confidence interval) under both protocols are 1.5 $(1.3, 1.8)$, significantly higher than the benchmark's 0.93 $(0.86, 1.00)$. The mean (95% confidence interval) certainty equivalent from the rolling protocol is: 497 $(291, 688)$ basis points per month. The optimal portfolios under both protocols have higher means and scales than both benchmark portfolios. Table 2 shows that both benchmarks are significantly negatively skewed: the 95% confidence interval on the value-weighted benchmark's S_4 (SKEW) is $(-11.6$, -0.4). Both benchmarks also have significantly fatter tails than a Gaussian distribution. The 95% confidence interval on the value-weighted benchmark's K_3 (KURT) is $(17, 35)$. The dynamic optimal portfolios from both rolling and updating, in contrast, are symmetric and not significantly more leptokurtic than a Gaussian distribution. The rolling protocol generates a portfolio with higher scale and sampling variation, and lower minimum returns than the updating protocol.

[11] The joint results on beta and residual volatility during this era are consistent with the findings in Liu, Stambaugh, and Yuan (2018). It is clear that prior to 1999, there is a relationship between variance and future expected returns. When the characteristic set contains both beta and residual volatility, beta becomes desirable and high residual volatility stocks appear undesirable, ceteris paribus. When the characteristic set includes beta but not residual volatility, then the sign on beta's θ coefficient becomes negative.

Table 2

Sampling properties of out-of-sample Portfolio Performance Statistics: 108-month out-of-sample period, 1990–1998

A. Benchmark portfolios

	VWI							EWI						
Statistic	Mean	Std Dev	2.5%ile	25%ile	Median	75%ile	97.5%ile	Mean	Std Dev	2.5%ile	25%ile	Median	75%ile	97.5%ile
$\mathcal{CE}_2$	128.8	4.0	121.0	126.1	128.8	131.5	136.6	109.4	2.3	104.8	107.9	109.4	110.9	113.9
$E(r)$	144.2	4.0	136.2	141.5	144.2	146.9	152.0	128.6	2.3	124.1	127.1	128.6	130.1	133.1
σ	388.6	4.6	379.6	385.5	388.6	391.7	397.7	430.0	2.3	425.1	428.3	430.0	431.6	434.9
Median	167.4	11.6	144.9	159.3	167.6	175.3	189.7	187.5	9.8	168.1	180.0	187.6	194.1	206.5
IQR	469.9	22.4	426.6	454.5	469.8	484.9	514.7	508.7	17.5	474.6	496.7	508.7	520.5	543.3
MIN	− 1,482.9	59.1	− 1,602.5	− 1,521.9	− 1,481.9	− 1,442.2	− 1,368.6	− 1,747.3	28.2	− 1,803.2	− 1,766.3	− 1,747.3	− 1,728.2	− 1,691.9
SKEW	− 6.0	2.9	− 11.6	− 8.0	− 6.0	− 4.0	− 0.4	− 13.7	2.2	− 18.0	− 15.2	− 13.7	− 12.2	− 9.3
KURT	26.4	4.6	17.4	23.3	26.4	29.4	35.6	35.2	2.5	30.3	33.5	35.2	36.9	40.2
SR	0.9337	0.0364	0.8627	0.9095	0.9337	0.9581	1.0051	0.7158	0.0187	0.6787	0.7034	0.7159	0.7284	0.7520

B. Dynamic PPP

	Dyn. opt. updating							Dyn. opt. rolling						
Statistic	Mean	Std Dev	2.5%ile	25%ile	Median	75%ile	97.5%ile	Mean	Std Dev	2.5%ile	25%ile	Median	75%ile	97.5%ile
$\mathcal{CE}_2$	427.7	50.9	328.6	393.6	428.1	461.1	529.0	497.2	246.8	291.1	444.1	507.7	570.1	688.4
$E(r)$	587.9	54.4	486.4	550.1	586.1	623.1	700.9	900.1	93.5	725.2	835.7	897.0	960.5	1,094.8
σ	1,254.8	88.6	1,090.0	1,193.7	1,251.6	1,311.1	1,441.2	1,952.8	129.8	1,714.8	1,862.7	1,947.4	2,035.4	2,221.6
Median	593.3	85.2	431.2	535.3	591.7	649.8	767.1	878.5	136.0	617.7	786.6	877.4	966.9	1,153.3
IQR	1,495.0	157.8	1,200.8	1,385.5	1,490.0	1,598.4	1,822.1	2,463.4	247.0	1,995.5	2,294.0	2,456.5	2,627.1	2,961.3
MIN	− 3,887.9	705.9	− 5,336.6	− 4,344.9	− 3,871.2	− 3,399.8	− 2,567.2	− 5,145.8	1,160.3	− 7,575.9	− 5,901.6	− 5,063.5	− 4,286.7	− 3,177.4
SKEW	− 0.4	5.3	− 10.9	− 4.1	− 0.4	3.2	10.2	1.1	5.7	− 10.0	− 2.7	1.0	4.9	12.5
KURT	24.1	13.4	− 1.2	14.9	23.9	32.9	50.7	11.8	14.2	− 15.2	1.9	11.4	21.2	40.8
SR	1.5281	0.1383	1.2626	1.4353	1.5266	1.6191	1.8050	1.5400	0.1391	1.2741	1.4451	1.5369	1.6333	1.8157

Sampling properties of dynamic optimal PPPs. Portfolio characteristic-tilts from the best out-of-sampling performer over the relevant preceding period (shown in Table 1) each year. $\mathcal{CE}_2$ is the certainty equivalent return in basis points per month for a power utility investor with coefficient of relative risk aversion $(\gamma) = 2$. $E(r)$, σ, Median, IQR, and MIN are the mean monthly return, the standard deviation of monthly returns, the median monthly return, the interquartile range of monthly returns, and the minimum monthly return—all expressed in basis points per month. SKEW and KURT are the return skewness and kurtosis measures, and SR is the Sharpe ratio. Results are for the first 9-year out-of-sample subperiod (1990–1998).

The left-hand panel of Figure 2 contrasts the optimal dynamic portfolio from the updating protocol with the value-weighted benchmark. The difference in scale is apparent, as is the fact that the distribution of the optimal portfolio seems shifted to the right of the benchmark (higher mean and median returns).

Comparing characteristic-tilted portfolios and benchmarks in the metric of certainty equivalent returns makes no assumptions about the sources of systematic risk, or the factor structure of returns. Since the dynamic PPP dominates the benchmarks, we next examine the relationship between these portfolios' returns and the Fama-French six-factor model of returns. Table 3, panel A provides information on how characteristics achieved higher utility in the pre-1999 era. This shows the projection of optimal portfolio returns minus the monthly riskless rate on the six Fama-French factors: the excess return on the value-weighted U.S. stock market (Mkt), the value factor (HML), the (small) size factor (SMB), the momentum factor (MOM), the profitability factor (RMW), and the investment factor (CMA), obtained from Professor Kenneth French's website at Dartmouth. The loading on the market factor is significantly negative under both protocols—that is, characteristics tilt the portfolio weights so that it is short the overall stock market factor. The power utility investor with $\gamma = 2$ instead seeks positive exposure to the value factor (HML), small stock factor (SMB), momentum factor (MOM), and the profitability factor (RMW). This portfolio has a significant negative loading on the conservative investment factor (CMA). The mean (95% confidence interval) portfolio alpha from the updating protocol is 264 $(157, 371)$ basis points per month. The analogous sampling statistics on alpha from the rolling protocol's optimal portfolio are 387 $(210, 578)$ basis points per month.

In this subperiod the characteristics shift the portfolio to lie outside the span of these six factors. Panel B of Table 3 decomposes the mean and variance of the updating protocol dynamic PPP's portfolio returns in excess of the risk-free rate within the space of the six-factor model. The orthogonal variance for each sample is the ratio of the residual variance from this regression to the portfolio variance. Variances attributed to the factors are the squared factor beta in the sample times the factor's variance. The variance values in the table do not sum to 100 in any bootstrap sample because the factors are not orthogonal. The largest pairwise correlations amongst the six factors in this first subperiod are: 77% between HML and CMA, -63% between the market and CMA, and -42% between SMB and RMW. On average (95% confidence interval), 53% $(45\%, 62\%)$ of the variance in excess returns is not spanned by the six-factor model. And 48% $(33\%, 61\%)$ of the portfolio mean excess return comes in the form of alpha from this six-factor model. Within the factor span, momentum accounts for an average (95% confidence interval) of 48% $(38\%, 60\%)$ of the optimal portfolio's expected excess return and 34% $(25\%, 43\%)$ of its variance. HML is the third-largest source of this portfolio's returns. HML accounts for 14% $(11\%, 19\%)$ of the optimal portfolio's expected excess return and 38% $(23\%, 54\%)$ of its variance. Table 3 panel B shows that on average (95% confidence interval), 2% $(0\%, 4\%)$ of the optimal portfolio's variance comes from exposure to the market factor.

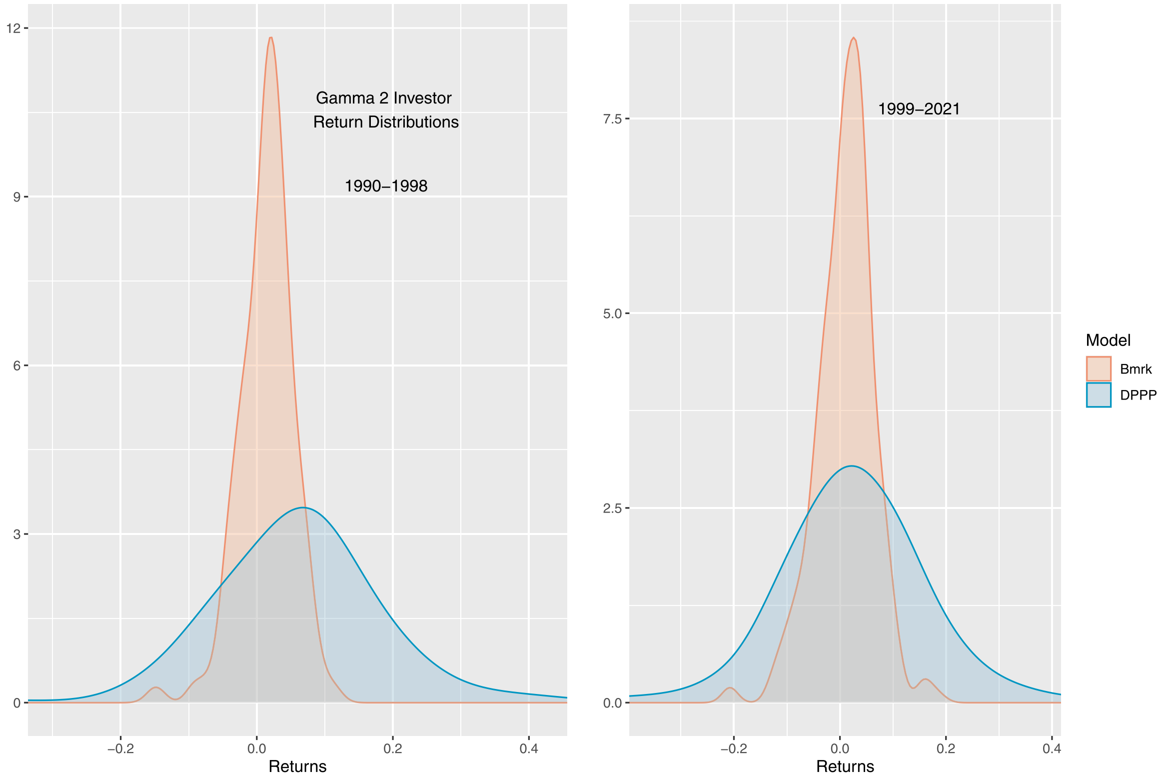

Figure 2
Portfolio return densities in the two subperiods
"DPPP" is the optimal dynamic parametric portfolio under the updating protocol—selected at the beginning of each year. "Bmrk" is the preferred benchmark in the subperiod, which for this power utility investor with coefficient of relative risk aversion, γ, = 2 are the value-weighted portfolio in the first subperiod and the equally-weighted portfolio in the second subperiod.

Table 3

Out-of-sample six-factor Fama-French regressions: $r_{i,t} - r_f = \alpha + \beta_1 \cdot (R_{m,t} - r_f) + \beta_2 \cdot \text{HML} + \beta_3 \cdot \text{SMB} + \beta_4 \cdot \text{MOM} + \beta_5 \cdot \text{RMW} + \beta_6 \cdot \text{CMA} + \epsilon_{i,t}$, for power utility investor with coefficient of relative risk aversion, γ=2; monthly returns; α in basis points per month

A. Subperiod 1: 1990–1998

	Updating protocol							Rolling protocol						
Coefficient	Mean	Std Dev	2.5%ile	25%ile	Median	75%ile	97.5%ile	Mean	Std Dev	2.5%ile	25%ile	Median	75%ile	97.5%ile
α / Orthog.	263.75	54.97	156.53	226.66	263.04	299.71	371.25	387.20	94.63	209.65	323.21	385.11	450.49	577.66
Mkt	−0.36	0.17	−0.69	−0.47	−0.36	−0.25	−0.04	−0.51	0.27	−1.05	−0.69	−0.51	−0.33	−0.02
HML	3.16	0.43	2.35	2.86	3.15	3.44	4.02	5.83	0.66	4.59	5.39	5.82	6.27	7.14
SMB	1.68	0.28	1.12	1.48	1.68	1.87	2.23	0.78	0.43	−0.06	0.50	0.78	1.06	1.64
MOM	2.62	0.25	2.14	2.44	2.61	2.78	3.13	3.80	0.39	3.04	3.53	3.80	4.06	4.58
RMW	0.93	0.41	0.14	0.65	0.92	1.20	1.75	1.42	0.74	−0.03	0.93	1.41	1.92	2.89
CMA	−1.67	0.53	−2.73	−2.01	−1.66	−1.30	−0.66	−4.08	0.86	−5.81	−4.65	−4.07	−3.50	−2.42

B. Subperiod 1: 1990–1998 Updating Protocol: Decompositions

	% of Portfolio Mean due to:							% of Portfolio Variance due to:						
Coefficient	Mean	Std Dev	2.5%ile	25%ile	Median	75%ile	97.5%ile	Mean	Std Dev	2.5%ile	25%ile	Median	75%ile	97.5%ile
α / Orthog.	47.86	7.07	32.94	43.41	48.22	52.74	60.76	53.16	4.26	45.09	50.24	53.12	56.00	61.72
Mkt	−6.64	3.15	−13.04	−8.71	−6.56	−4.50	−0.65	1.52	1.17	0.03	0.63	1.27	2.17	4.42
HML	14.27	2.02	10.63	12.90	14.18	15.54	18.52	37.75	7.77	23.31	32.39	37.46	42.86	53.65
SMB	−7.90	1.62	−11.23	−8.94	−7.85	−6.79	−4.90	13.45	4.34	5.74	10.36	13.16	16.25	22.65

(continued)

Table 3
Continued

B. Subperiod 1: 1990–1998 Updating Protocol: Decompositions

	% of Portfolio Mean due to:							% of Portfolio Variance due to:						
Coefficient	Mean	Std Dev	2.5%ile	25%ile	Median	75%ile	97.5%ile	Mean	Std Dev	2.5%ile	25%ile	Median	75%ile	97.5%ile
MOM	48.11	5.57	38.05	44.32	47.83	51.57	59.98	34.07	4.65	25.09	30.97	34.05	37.17	43.41
RMW	7.35	3.29	1.10	5.12	7.26	9.49	14.02	1.22	0.92	0.04	0.52	1.03	1.72	3.51
CMA	−3.06	0.97	−5.01	−3.68	−3.05	−2.40	−1.23	6.61	3.73	1.03	3.84	6.11	8.75	15.32

C. Subperiod 2: 1999–2021

	Updating protocol							Rolling protocol						
Coefficient	Mean	Std Dev	2.5%ile	25%ile	Median	75%ile	97.5%ile	Mean	Std Dev	2.5%ile	25%ile	Median	75%ile	97.5%ile
α / Orthog.	80.86	34.66	14.47	56.94	80.18	104.21	150.37	99.24	30.89	40.38	78.48	98.04	119.61	161.97
Mkt	0.91	0.11	0.70	0.84	0.91	0.98	1.12	1.27	0.09	1.10	1.20	1.27	1.33	1.47
HML	1.38	0.16	1.07	1.27	1.38	1.49	1.70	0.74	0.14	0.47	0.65	0.73	0.83	1.02
SMB	0.62	0.22	0.18	0.47	0.62	0.77	1.04	−1.14	0.26	−1.69	−1.31	−1.13	−0.95	−0.67
MOM	1.01	0.15	0.71	0.91	1.01	1.12	1.32	0.50	0.14	0.23	0.41	0.50	0.59	0.78
RMW	1.26	0.22	0.83	1.11	1.26	1.41	1.70	0.01	0.31	−0.59	−0.19	0.02	0.22	0.61
CMA	−0.22	0.25	−0.70	−0.39	−0.22	−0.05	0.20	−0.13	0.25	−0.62	−0.29	−0.12	0.04	0.35

2.2.2 Subperiod 2 (1999–2021). Table 4 reports the properties of the dynamically optimized portfolios in the second-stage out-of-sample subperiod, 1999–2021. The right-hand panel in Figure 2 shows the return densities of the optimal portfolio from the updating protocol and the equally-weighted portfolio of sample stocks in this period. After 1998, both benchmark portfolios dominate the regularized dynamic parametric portfolio. The mean (95% confidence interval) certainty equivalent from the updating dynamic PPP is $-2\,(-110\,,83)$ basis points per month, whereas these statistics are 77 $(73, 80)$ basis points per month for the equally-weighted benchmark and 60 $(53, 67)$ basis points per month for the value-weighted benchmark. The dynamic parametric portfolios under both protocols have significantly higher mean returns than the benchmarks. The Sharpe ratio from the updating protocol is not statistically different from the benchmarks' Sharpe ratios. However, the equally-weighted benchmark's Sharpe ratio is significantly larger than that generated by the rolling protocol's dynamic PPP.

As in the first subperiod, characteristic-tilts generate portfolios that are not negatively skewed—unlike the benchmarks. However, in this subperiod the dynamic PPP is dominated by the benchmarks for several reasons. First, characteristics are used to increase the scale of the distribution (measured by the interquartile range) by some 3.4 fold in both subperiods. The portfolio median return is only 1.9 times higher than the benchmark in this subperiod, whereas this ratio is 3.5 in the first subperiod. Another feature of the characteristic-based portfolios that changes from the first to the second subperiod is heightened kurtosis—manifest in the large negative returns. This is especially true for the rolling protocol under which at least one monthly return is less than -100% in more than 25% of the bootstrap samples, and less than -87% in more than 75% of these samples. Even though the power utility investor with $\gamma = 2$ is relatively risk tolerant, losing 100% results in a certainty equivalent return of that same magnitude.

Table 3, panel C shows that the optimal characteristic-based portfolio from the updating protocol has a mean (95% confidence interval) alpha of 81 $(14, 150)$ basis points per month, which is significantly positive. This highlights the possibility for discrepancies between the performance metrics, as this portfolio's Sharpe ratio is not significantly different from the benchmark. Because it results in a negative certainty equivalent return in more than 25% of the bootstrap samples, we consider it dominated in the metric of expected utility. This discrepancy is even more pronounced by the optimal dynamic parametric portfolio from the rolling protocol. This portfolio's alpha is larger (although not significantly so) than that from the updating protocol. Yet this rule produces a portfolio whose certainty equivalent is less than -21% (per month) in more than half of the bootstrap samples. Comparing the two periods, the biggest effect on the portfolio mean is the drop in the loading on momentum, whose mean drops from 2.6 to 1.0. The mean return on this factor dropped from 100 to 26 basis points per month.[12]

[12] This phenomenon likely related to the momentum crash of 2009, documented by Barroso and Santa-Clara (2015a) and Daniel and Moskowitz (2016).

Table 4

Sampling properties of out-of-sample portfolio performance statistics: 276-month out-of-sample period, 1999–2021

A. Benchmark portfolios

	VWI							EWI						
Statistic	Mean	Std Dev	2.5%ile	25%ile	Median	75%ile	97.5%ile	Mean	Std Dev	2.5%ile	25%ile	Median	75%ile	97.5%ile
$\mathcal{CE}_2$	60.0	3.5	53.1	57.6	60.0	62.4	66.9	76.6	1.7	73.4	75.4	76.6	77.7	79.8
$E(r)$	79.2	3.5	72.4	76.8	79.2	81.6	85.9	106.8	1.7	103.6	105.6	106.8	107.9	110.0
σ	433.3	4.2	425.2	430.4	433.3	436.3	441.6	542.7	2.1	538.6	541.3	542.7	544.1	546.8
Median	124.0	8.0	107.9	118.6	124.1	129.4	139.4	141.9	6.7	128.8	137.4	141.9	146.4	155.3
IQR	511.2	13.9	484.6	501.8	510.9	520.6	538.8	655.0	12.1	612.0	626.8	635.0	642.9	659.3
MIN	− 1,667.4	78.6	− 1,819.4	− 1,720.6	− 1,667.6	− 1,613.6	− 1,514.3	− 2119.7	42.4	− 2,206.5	− 2,147.9	− 2,118.6	− 2,090.3	− 2,039.7
SKEW	− 10.3	1.8	− 13.9	− 11.5	− 10.4	− 9.1	− 6.8	− 6.5	1.2	− 8.9	− 7.3	− 6.5	− 5.6	− 4.1
KURT	27.9	3.8	20.6	25.3	27.9	30.4	35.4	32.5	1.5	29.5	31.5	32.5	33.5	35.5
SR	0.5245	0.0286	0.4687	0.5052	0.5246	0.5436	0.5807	0.5959	0.0105	0.5756	0.5887	0.5957	0.6029	0.6165

B. Dynamic PPP

	Dyn. opt. updating							Dyn. opt. rolling						
Statistic	Mean	Std Dev	2.5%ile	25%ile	Median	75%ile	97.5%ile	Mean	Std Dev	2.5%ile	25%ile	Median	75%ile	97.5%ile
$\mathcal{CE}_2$	− 1.5	172.3	− 110.2	− 24.9	5.9	33.5	82.9	− 5,017.2	4,7619.8	− 10,000	− 10,000	− 2,099.6	− 222.1	4.1
$E(r)$	242.7	32.8	180.1	220.1	242.2	264.3	309.4	175.2	28.1	121.4	156.3	174.7	193.0	232.9
σ	1,472.3	72.6	1,339.3	1,421.4	1,469.2	1,519.9	1,622.7	1,370.2	144.2	1,120.7	1,269.2	1,356.7	1,459.4	1,685.5
Median	233.0	52.2	133.8	196.7	232.7	267.9	337.9	149.2	22.8	107.0	132.9	148.5	164.9	194.5
IQR	1,718.9	113.3	1,510.8	1,639.5	1,714.3	1,792.7	1,954.6	748.9	37.6	680.4	722.2	747.6	773.7	826.2
MIN	− 5,490.7	933.4	− 7,756.8	− 5,958.2	− 5,352.3	− 4,856.6	− 4,072.5	− 10,082.7	1,862.8	− 14,309.7	− 11,127.4	− 9,861.2	− 8,787.2	− 7,071.3
SKEW	0.6	3.2	− 5.7	− 1.5	0.7	2.8	6.8	1.9	2.1	− 2.3	0.5	1.9	3.3	6.0
KURT	40.2	11.5	18.6	32.3	39.7	47.7	63.2	209.8	23.5	164.1	193.6	209.6	225.4	256.9
SR	0.5415	0.0750	0.3971	0.4895	0.5414	0.5931	0.6902	0.4130	0.0726	0.2717	0.3641	0.4127	0.4617	0.5566

Sampling properties of dynamic optimal PPPs. Portfolio characteristic-tilts from the best out-of-sample performer over the relevant preceding period (shown in Table 1) each year. $\mathcal{CE}_2$ is the certainty equivalent return in basis points per month for a power utility investor with coefficient of relative risk aversion $(\gamma) = 2$. $E(r)$, σ, Median, IQR, and MIN are the mean monthly return, the standard deviation of monthly returns, the median monthly return, the interquartile range of monthly returns, and the minimum monthly return—all expressed in basis points per month. SKEW and KURT are the return skewness and kurtosis measures, and SR is the Sharpe ratio. Results are for the second 23-year out-of-sample subperiod (1999–2021).

The second-largest effect is the drop in alpha—the mean alpha drops by 183 basis points per month.

These effects are consistent with the properties of the θ coefficients under the rolling protocol, shown in Figure 1. The optimal portfolios shift toward the market factor in this period. The mean (95% confidence interval) loading on the market portfolio in this period is 0.9 $(0.7, 1.1)$ under the updating protocol and 1.3 $(1.1, 1.5)$ under the rolling protocol.

2.3 Updating versus rolling

In our empirical design we do not allow the investor to choose between rolling and updating, since the out-of-sample periods used for identifying the ex post optimal configurations are not the same across the two protocols. Instead, as econometricians we compare results across protocols to make inferences about the nature of the data-generating process.

Table 1 shows that the updating protocol does not provide any warning that characteristics' predictive value has vanished. Instead, the ex post optimal model appears temporally stable. The large gains over the first 15 years of the configuration-selection period disguise the fact that the PPPs underperform benchmarks in the more recent years. For example, were we to evaluate the model using the ex post optimal portfolio ending in 2018 (Table 1), we would infer that the optimal value for λ is 1, and the optimal characteristic set comprises: momentum, log size, beta, average same-month return, and residual volatility. Indeed, over the out-of-sample period from 1975–2018 this portfolio dominates the benchmarks. However, this masks the fact that this portfolio's certainty equivalent mean (95% confidence interval) over the preceding 180 months is -23 $(-113, 45)$ basis points per month. The salutary historical performance over the 44-year out-of-sample period trades off the positive effects of characteristic-based tilting in the years 1975–1998 against the negative effects in the latter years. This underscores the issues raised in Martin and Nagel (2022). The flexibility to choose the characteristic set and the hyperparameter, λ, expands the dimensionality of the model space. The results of using an ex post optimal model configuration (on out-of-sample returns) must be analyzed in a subsequent (truly) out-of-sample period to evaluate the model.

Table 1 suggests that some of the usefulness of λ as a regularization parameter may derive from its capacity to detect a structural break. However, this is primarily true under the rolling protocol. Once there is a year in which the out-of-sample results deteriorate there is a jump in λ, as the algorithm scales back its reliance on characteristics. When 2000 replaces 1985 in the out-of-sample validation period, λ jumps from 1 to 5. This increase in λ has a large impact on the out-of-sample certainty equivalent returns. For contrast, the 1 percentile (95% confidence interval) on the certainty equivalent return from the configuration that was optimal in the preceding year, with $\lambda = 1$, is $-10{,}000$ $(-10{,}000, 398.4)$.

Since Table 2 describes a noisy, structurally stable environment, the updating protocol produces portfolios with higher 2.5 percentile certainty equivalents than rolling for all three investors. However the rolling protocol produces higher 97.5 percentile certainty equivalents than updating. In a structurally stable environment more data is strictly preferred. Table 1 shows that the two protocols have very similar ex post optimal configurations (λ and characteristic sets) through 1999.

2.4 Increasing risk aversion

Figures IA-1 to IA-6 show the confidence intervals of the ex post optimal θ values from the rolling protocol preceding each year in the second-stage out-of-sample period, for each of the six characteristics for three power utility investors, with coefficients of relative risk-aversion: 2, 5, and 8. The values for the least risk-averse investor correspond to those reported in Figure 1. The figures show that as risk-aversion increases the optimal effect of characteristics on portfolio weights decreases in absolute value. Further, the apparent appeal of characteristic-tilt appears to vanish more quickly for the more risk-averse investors. For example, the ex post optimal out-of-sample portfolio puts a significant positive weight on average same month return for the first 20 years in the second-stage out-of-sample period, for the most risk-tolerant investor—although there is a significant drop in the coefficient after 2000 (for the 2001 out-of-sample portfolio). However, this coefficient is zero for the most risk-averse investor in years 12–26 and 28–32.

Tables IA-2 and IA-3 provide the ex post optimal configurations prior to each year in the second-stage out-of-sample period, from both protocols (analogous to Table 1) for the power utility investors with $\gamma = 5$ and $\gamma = 8$, respectively. Both power utility investors optimally set $\lambda = 0$ for most of the first subsample, confirming Brandt, Santa-Clara, and Valkanov's conjecture that their algorithm has much less estimation risk than mean-variance optimization. Contrasting this with the importance of $\lambda > 0$ for the most risk-tolerant of these three investors provides additional evidence that estimation risk (the tendency to overfit) is linked to the return variance. The power utility investor with $\gamma = 2$ is tolerant enough of variance (and kurtosis) so that maximizing this utility function directly on the data accepts positions with attractive in-sample expected return and skew that result from noise in the conditional return generating process. In this case using $\lambda = 1$ helps to separate the predictability from the noise.

These tables also demonstrate similar evidence from the rolling protocol, as in Table 1, that the fit of the model starts to deteriorate around 1999. The equally-weighted benchmark is the ex post optimal portfolio (i.e., the 1 percentile of its certainty equivalent sampling distribution is greater than all 882 configurations) in 2009–2015 for the power utility investor with $\gamma = 5$. Similarly, the power utility investor chooses either the value-weighted benchmark or the equally-weighted benchmark above all 882 optimized portfolios prior to the years: 2002, 2005, and 2009–2015.

Table IA-4 is analogous to Table 2. It shows that in the first true out-of-sample period (1990–1998), the dynamic PPP portfolios under the updating protocol dominate the benchmark portfolios for both of these more risk-averse power utility investors. However, unlike the case for our most risk-tolerant investor, the optimal dynamic PPP portfolios from the rolling protocol do not dominate the benchmarks in this period for these investors. These results jointly lend credence to the hypothesis that the conditional return-generating process was largely stable over the 1955–1998 period, as using more information reduces estimation risk. This table also shows that the optimal dynamic portfolios from the updating protocol are neither negatively skewed nor more leptokurtic than the Gaussian distribution.

The left-hand panels in Figures IA-7 and IA-8 contrast the return distributions of the optimal portfolio from the updating protocol with the preferred benchmark, which is the value-weighted benchmark in both cases in the first subperiod. Figure IA-8 is especially revealing as our most risk-averse investor's optimal portfolio has a similar scale as the benchmark. The mean (95% confidence interval) interquartile range of the optimal dynamic portfolio is 702 $(576, 840)$ basis points per month compared with the value-weighted benchmark's 470 $(427, 515)$ basis points per month. This figure's left panel shows that the optimal portfolio has more significant mass above a 5% monthly return than the benchmark, while having similar left-tail properties in the first subperiod. Both dynamic optimal portfolios have similar Sharpe ratios to that in Table 2 for our most risk-tolerant investor. Unlike the certainty equivalent, the Sharpe ratios from both protocols are very similar—both are significantly larger than the benchmarks'.

Table IA-5 reports the out-of-sample properties of the dynamic optimal portfolios for these two investors in our second-stage out-of-sample subperiod, 1999–2021. The mean certainty equivalent is negative for both investors' dynamic optimal portfolios under both protocols in this subperiod. The updating portfolios' Sharpe ratios are not significantly different from the benchmarks'; however, both investors' optimal dynamic portfolios' Sharpe ratios under the rolling protocol are significantly lower than the equally-weighted benchmarks'.

Table IA-6 contrasts the dynamic PPP with the benchmarks for the entire 384 month out-of-sample period (1990–2021). The optimal portfolios for our focal investor (whose $\gamma = 2$) is the equally-weighted benchmark, and the optimal portfolio for the two more risk-averse investors is the value-weighted benchmark. As in the second subperiod, the dynamic PPP is less attractive in the expected utility metric because of its additional kurtosis. The two more risk-averse investors' optimal portfolios obtained using the updating protocol have significantly higher Sharpe ratios than the benchmarks. Unlike the utility function, the Sharpe ratio does not over-weight the smaller minimum returns. The mean (95% confidence interval) of the value-weighted benchmark's minimum monthly return over this period is -17% (-18%, -15%) per month. These statistics for our most risk-averse investor's dynamic PPP from the updating protocol are -33% (-44%, -25%) per month. These distributions are compared graphically in the right-hand panels of Figures IA-7 and IA-8. Both of these graphs show that the left tail of the dynamic PPP's return distribution dominates that of the preferred benchmark.

Tables IA-7 and IA-8 report the relationships between these two more risk-averse investors' optimal portfolios and the six Fama-French factors for all three out-of-sample periods, the first and second subperiods and the full 32-year period. Consistent with the analysis of these portfolios above, alpha decreases in γ. In the first subperiod our most risk-averse investor's optimal portfolio has mean (95% confidence interval) alpha of 104 $(60, 149)$ basis points per month, (Table IA-8). Table IA-9 provides the sampling distributions of the power utility investor with $\gamma = 2$ dynamic optimal portfolio return projection on the Fama-French six factors for the full 32-year out-of-sample period. Table IA-10 reports the mean and variance decompositions for the two more risk-averse investors' optimal portfolios in the first subperiod (analogous to Table 3, panel B). During this subperiod, when the characteristics provide useful information about future returns, the variance decompositions are constant across risk aversion. For all three investors the six factors account for 40% to 55% (95% confidence intervals) of the optimal portfolio's return variance. Similarly, HML and MOM are the largest sources of return variance for all three portfolios amongst the six factors for all three investors.

In the period when characteristics were efficacious, the optimal portfolios' scales decrease in risk aversion. The interquartile range of returns for our most risk-tolerant investor has a mean (95% confidence interval) of 1,495 $(1,201, 1,822)$ basis points per month; the analogous statistics for our most risk-averse investor's dynamic optimal portfolio are 702 $(576, 840)$ basis points per month. However the sources of variance in the span of the six Fama-French factors are flat in γ. This result is fully consistent with the notion that sources of predictable excess return require exposure to non-diversifiable variance, as discussed by Kozak, Nagel, and Santosh (2020). Since all three investors tilt their optimal portfolios toward stocks that have done relatively well in the same month over the past five years, this result is consistent with Keloharju, Linnainmaa, and Nyberg's (2016) hypothesis that there are many small seasonal (month-of-the-year) factors, and return in the month serves as an instrument for exposure to these factors.

3. Conclusions

We explore the nature of estimation risk in conjunction with Brandt, Santa-Clara, and Valkanov's (2009) parametric portfolio policy. We use a bootstrapped out-of-sample max-min criterion to select the optimal portfolio configuration at the beginning of each year in a second-stage (optimized) out-of-sample period. To gain insight into the interactions between portfolio optimization and estimation risk, we introduce a new form of regularization. Rather than penalizing the parameter space we introduce a hyperparameter that can increase the curvature of the loss function used to estimate portfolio weights. For power utility investors with moderate to high risk aversion (coefficients of relative risk aversion of 5 and 8), Brandt, Santa-Clara, and Valkanov's (2009) PPP algorithm afforded a way to exploit this predictability without significant estimation error. The most risk-tolerant power utility investor we consider, with coefficient of relative risk aversion of two, experiences much higher estimation risk with the parametric portfolio policy. This investor reduces overfitting by optimizing a power utility function with a higher coefficient of relative risk aversion than her own in sample. This loss function increases the shadow cost of variance in terms of expected return. It works because estimation error increases in portfolio variance.

Our results suggest that measurable characteristics did have economically and statistically significant predictive content for portfolio construction prior to the year 1999. That result considers all moments of the predictive distribution and is not specific to expected return, alpha, or Sharpe ratio. Since during this period (the twentieth century) optimal portfolios' characteristic-tilts diminished in risk aversion, we do not infer that any of the opportunities afforded by conditioning on characteristics represented a free lunch. Rather there were dimensions wherein the tradeoffs between mean (and skewness) and variance (and kurtosis) were more attractive than those afforded by the market portfolio. Two of those dimensions appear to be the well-known momentum and value factors. Optimal portfolio returns were virtually orthogonal to the market factor and one-half of their return variance came from outside the span of the Fama-French six-factor model.

Characteristics' predictive efficacy for portfolio optimization has vanished starting in 1999. The market portfolio dominates all of the dynamically optimal parametric portfolios over the 1999–2021 period. This finding is consistent with the literature, which contains several non-mutually exclusive hypotheses. Martin and Nagel (2022) show that in a complex economy in which agents learn about predictive relationships, econometricians should expect to find in-sample predictability that vanishes out of sample. McLean and Pontiff (2016) suggest that investors adapt to academic research. Green, Hand, and Zhang (2017) note that the twenty-first century has seen new regulations and technological advances that serve to reduce trading frictions. This, in turn, allows investors to more fully exploit predictive relationships.

Code availability

The replication code is available in the Harvard Dataverse at https://doi.org/10.7910/DVN/LK4DCN.

Appendix

This appendix provides a pseudo code for our bootstrap dynamically regularized out-of-sample empirical design, with both updating and rolling sample protocols. Each year we optimize expected utility in sample to construct the next year's out-of-sample returns. Once we have at least 180 months of out-of-sample returns, we select the optimal configuration with numerical max-min of out-of-sample certainty equivalent returns. The optimal dynamically regularized policy—or second out-of-sample stage—has this configuration's out-of-sample returns in the next year. As in the text, we use the following notation: $y = 1, \ldots, 62$ references the number of each year in our sample. Uppercase Y is the year: $Y_1 = 1960$, $Y_{15} = 1974$, $Y_{30} = 1989$, and $Y_{62} = 2021$.

FOR each year, $y = 15, 16, \ldots, 29$:

  FOR each configuration, $c = 1, 2, \ldots, 882$:

    FOR each bootstrap sample, $b = 1, 2, \ldots, 10,000$:

      Form out-of-sample returns: Maximize (1) over θ—using $[Y_s, Y_y]$.* These $\widehat{\theta}_{y,c,b}$ are used to construct the out-of-sample portfolio returns in the year Y_{y+1}, for configuration c and bootstrap sample b, respectively.

    END FOR

  END FOR

END FOR

FOR each year, $y = 30, 31, \ldots, 61$:

  Dynamic Regularization: Identify which of the 882 configurations' *out-of-sample portfolio returns* over years $[Y_\upsilon, Y_y]$ has the maximum 1 percentile value certainty equivalent return.** This optimal configuration, c_y^*, is reported by year (Y_{y+1}) in Table 1 for the power utility investor with coefficient of relative risk aversion of 2. Tables IA-2 and IA-3 report this optimal configuration by year for power utility investors with coefficients of relative risk aversion of 5 and 8, respectively.

  FOR each configuration, $c = 1, 2, \ldots, 882$:

    FOR each bootstrap sample, $b = 1, 2, \ldots, 10{,}000$:

      Form out-of-sample returns: Maximize (1) over θ—using $[Y_s, Y_y]$.* These $\widehat{\theta}_{y,c,b}$ are used to construct the out-of-sample portfolio returns in the year Y_{y+1}, for configuration c and bootstrap sample b, respectively.

    END FOR

  END FOR

  The bootstrap set of second-stage, out-of-sample dynamically regularized optimal portfolio returns for the 12 months in year $y + 1$ is: $\{r_{y+1,c_y^*}\}$.

END FOR

---

* Under the updating protocol, $Y_s \equiv Y_1 = 1960$, and $Y_s = Y_{y-14}$ under the rolling protocol.

** Under the updating protocol, $Y_\upsilon \equiv Y_{16} = 1975$, and $Y_\upsilon = Y_{y-14}$ under the rolling protocol.

References

Aït-Sahalia, Y., and M. Brandt. 2001. Variable selection for portfolio choice. *Journal of Finance* 56:1297–351.

Ao, M., Y. Li, and X. Zheng. 2019. Approaching mean-variance efficiency for large portfolios. *Review of Financial Studies* 32:2890–919.

Barroso, P., J. Reichenecker, M. Reichenecker, and F. Rouxelin. 2023. An international equity and currency optimisation with frictions. Working paper, Catloica-Lisbon School of Business and Economics.

Barroso, P., and P. Santa-Clara. 2015a. Momentum has its moments. *Journal of Financial Economics* 116:111–20.

———. 2015b. Beyond the carry trade: Optimal currency portfolios. *Journal of Financial and Quantitative Analysis* 50:1037–56.

Barroso, P., and K. Saxena. 2022. Lest we forget: Learn from out-of-sample forecast errors when optimizing portfolios. *Review of Financial Studies* 35:1222-78.

Bertsimas, D., V. Gupta, and N. Kallus. 2018. Data-driven robust optimization. *Mathematical Programming, Series A* 167:235–92.

Best, M., and R. Grauer. 1991. On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results. *Review of Financial Studies* 4:315–42.

Boguth, O., M. Carlson, A. Fisher, and M. Simutin. 2011. Conditional risk and performance evaluation: Volatility timing, overconditioning, and new estimates of momentum alphas. *Journal of Financial Economics* 102:363–89.

Brandt, M., P. Santa-Clara, and R. Valkanov. 2009. Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *Review of Financial Studies* 22:3411–47.

Daniel, K., and T. Moskowitz. 2016. Momentum crashes. *Journal of Financial Economics* 122:221–47.

DeMiguel, V., L. Garlappi, and R. Uppal. 2009. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies* 22:1915–53.

DeMiguel, V., A. Martín-Utrera, F. Nogales, and R. Uppal. 2020. A transaction-cost perspective on the multitude of firm characteristics. *Review of Financial Studies* 33:2180–222.

DeMiguel, V., Y. Plyakha, R. Uppal, and G. Vilkov. 2013. Improving portfolio selection using option-implied volatility and skewness. *Journal of Financial and Quantitative Analysis* 48:1813–45.

Faias, J., and P. Santa-Clara. 2017. Optimal option portfolio strategies: Deepening the puzzle of index option mispricing. *Journal of Financial and Quantitative Analysis* 52:277–303.

Freyberger, J., A. Neuhierl, and M. Weber. 2020. Dissecting characteristics nonparametrically. *Review of Financial Studies* 33:2326–77.

Gay, D. 1983. Subroutines for unconstrained minimization using a model/trust-region approach. *ACM Transactions on Mathematical Software* 9:503–24.

Gilboa, I., A. Postlewaite, and D. Schmeidler. 2008. Probability and uncertainty in economic modeling. *Journal of Economic Perspectives* 22:173–88.

Gilboa, I., and D. Schmeidler. 1989. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics* 18:141–53.

Green, J., J. Hand, and X. Zhang. 2017. The characteristics that provide independent information about average U.S. monthly stock returns. *Review of Financial Studies* 30:4389–436.

Hansen, L., and T. Sargent. 2008. *Robustness*, Princeton, NJ: Princeton University Press.

Heston, S., and R. Sadka. 2008. Seasonality in the cross-section of stock returns. *Journal of Financial Economics* 87:418–45.

Jagannathan, R., and T. Ma. 2003. Risk reduction in large portfolios: Why imposing the wrong constraint helps. *Journal of Finance* 58:1651–83.

Kadan, O., and F. Liu. 2014. Performance evaluation with high moments and disaster risk. *Journal of Financial Economics* 113:131–55.

Kan, R., and G. Zhou. 2007. Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis* 42:621–56.

Kapetanios, G. 2008. A bootstrap procedure for panel data sets with many cross-sectional units. *The Econometrics Journal* 11:377–95.

Keloharju, M., J. Linnainmaa, and P. Nyberg. 2016. Return seasonalities. *Journal of Finance* 71:1557–89.

Kim, T., and H. White. 2004. On more robust estimation of skewness and kurtosis. *Finance Research Letters* 1:56–73.

Kozak, S., S. Nagel, and S. Santosh. 2020. Shrinking the cross section. *Journal of Financial Economics* 135:271–92.

Kroencke, T., F. Schindler, and A. Schrimpf. 2014. International diversification benefits with foreign investment styles. *Review of Finance* 18:1847–83.

Lewellen, J. 2015 The cross section of expected stock returns. *Critical Finance Review* 4:1–44.

Lewellen, J., S. Nagel, and J. Shanken. 2010. A skeptical appraisal of asset pricing tests. *Journal of Financial Economics* 96:175–94.

Liu, J., R. Stambaugh, and Y. Yuan. 2018. Absolving beta of volatility's effects. *Journal of Financial Economics* 128:1–15.

MacKinlay, A., and L. Pástor. 2000. Asset pricing models: Implications for expected returns and portfolio selection. *Review of Financial Studies* 13:883–916.

Martin, I., and S. Nagel. 2022. Market efficiency in the age of big data. *Journal of Financial Economics* 145:154–77.

McLean, R., and J. Pontiff. 2016. Does academic research destroy stock return predictability? *Journal of Finance* 71:5–31.

Nagel, S. 2021. *Machine Learning in Asset Pricing*, Princeton, NJ: Princeton University Press.

Pástor, L. 2000. Portfolio selection and asset pricing models. *Journal of Finance* 60:179–223.

Pástor, L., and R. Stambaugh. 2000. Comparing asset pricing models: An investment perspective. *Journal of Financial Economics* 56:335–81.

Rahimian, H., and S. Mehrotra. 2022. Frameworks and results in distributionally robust optimization. *Open Journal of Mathematical Optimization* 3:1–85.

Review of Asset Pricing Studies, 2024, 14, 450–480
https://doi.org/10.1093/rapstu/raae006
Article