Internet Appendix (not for publication) for “Modeling Preference Heterogeneity Within and Across Behavioral Types: Evidence from a Real-world Betting Market”

Angie Andrikogiannopoulou and Filippos Papakonstantinou*

*Both authors are at King’s College London.

IA.1 Directed acyclic graph of the hierarchical model

In Figure IA.1, we present the Directed Acyclic Graph (DAG) representation of our hierarchical mixture model described in Section 1.2 of the paper. In this graph, squares represent quantities that are fixed or observed, e.g., prior parameters and data, while circles represent unknown model parameters that need to be estimated. We generally use weak priors, and we perform a prior sensitivity analysis which shows that our posteriors are quite robust to varying the priors.

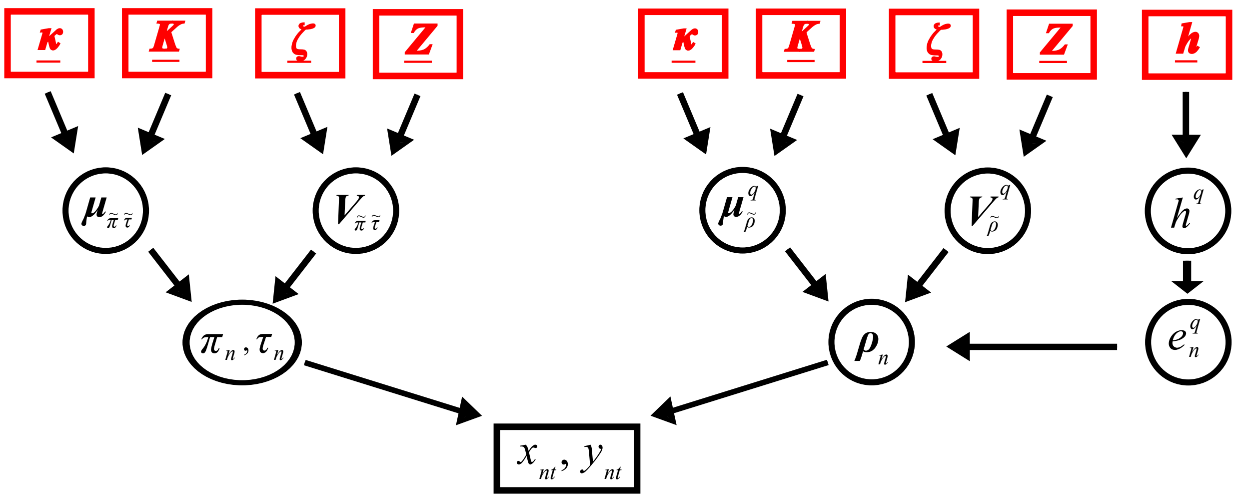


Figure IA.1: Representation of the hierarchical mixture model as a directed acyclic graph.

IA.2 Bet types and their lottery representation

All of the bets available in the sportsbook are combinations of various types of *accumulators*, where an accumulator of type $k \geq 1$ is a bet that involves k individual wagers and pays money only if all k wagers win.[1] An accumulator of type k can be represented as a lottery with two possible prizes: a positive prize that equals the stake times the product of the gross returns associated with each of the k selections minus one, and a negative prize with absolute value equal to the stake.[2] Since only unrelated selections can be combined into accumulators of type $k > 1$, the probability of the positive prize in such a bet is simply the product of the probabilities associated with each of the k selections involved.

A bet involving $N \geq 1$ individual wagers may belong to one of two broad categories of bet types: it could be (a) *a permutation* bet, which includes all accumulators of type k that could be constructed from the N wagers, for *some* $k \leq N$, i.e., it includes $\binom{N}{k}$ accumulators of type k, or (b) a *full-cover* bet, which includes all accumulators of type k that could be constructed from these N wagers, for *all* $1 \leq k \leq N$ or $2 \leq k \leq N$, i.e., it includes $\sum_{1 \leq k \leq N} \binom{N}{k}$ or $\sum_{2 \leq k \leq N} \binom{N}{k}$ accumulators of type 1 or 2 through N. The amount staked on permutation and full-cover bets is divided evenly among the number of accumulator bets they involve. For example, a bet that stakes amount S and involves $N = 3$ wagers could belong to one of the following five bet types:

- a *singles* bet that stakes amount $\frac{S}{3}$ on each of the $\binom{3}{1} = 3$ accumulators of type 1, each of which corresponds to one of the wagers and pays money if that wager wins,
- a *doubles* bet that stakes amount $\frac{S}{3}$ on each of the $\binom{3}{2} = 3$ accumulators of type 2, each of which corresponds to one of the pairwise combinations of wagers and pays money if both wagers in the combination win,
- a *treble* bet that stakes amount S on the $\binom{3}{3} = 1$ accumulator of type 3, which pays money if all three wagers win,
- a *trixie* bet which stakes amount $\frac{S}{4}$ on each of the aforementioned double bets (3 accumulators of type 2) *and* the aforementioned treble bet (1 accumulator of type 3), or
- a *patent* bet that stakes amount $\frac{S}{7}$ on each of the aforementioned single bets (3 accumulators of type 1) *and* each of the bets in the aforementioned trixie bet (3 accumulators of type 2 and 1 accumulator of type 3).

Note that the singles, doubles, and treble bets belong to the category of permutation bet types, while the trixie and patent bets belong to the category of full-cover bet types.

Lottery representations of permutation and full-cover bets can be constructed by combining the lottery representations of the elemental accumulator bets. Several bet calculators are available on the web that help bettors compute the prizes and probabilities of their chosen bets following the rules described above.

For example, consider an individual who submits the following bets:

- He stakes €50 on a *single* bet consisting of one wager in which the outcomes' odds are 1.33 and 2.80,

[1]The simplest bet type is an accumulator of type 1. Accumulators of type 1, 2, and 3 are usually called single, double, and treble bets, respectively.

[2]For example, a bet that combines a wager to back Barcelona to draw Real Madrid at odds 4.00 with a wager to back Federer to beat Hartfield in the first set at odds 1.60 is an accumulator bet of type 2. This accumulator would have a gross (net) payoff equal to the amount of money staked times $4.00 \times 1.60 = 6.40$ ($6.40 - 1 = 5.40$) if both wagers won, and a gross (net) payoff equal to 0 (minus the stake) otherwise.

and the backed outcome's odds are 1.33.

- He stakes €30 on a *doubles* bet consisting of three wagers: (i) in the first wager, the outcomes' odds are 2.25 and 1.56, and the backed outcome's odds are 2.25, (ii) in the second wager, the outcomes' odds are 1.67 and 2.00, and the backed outcome's odds are 1.67, and (iii) in the third wager, the outcomes' odds are 1.10 and 4.60, and the backed outcome's odds are 1.10.
- He stakes €2 on a *double* bet consisting of two wagers: (i) in the first wager, the outcomes' odds are the outcomes' odds are 4.00 and 1.20, and the backed outcome's odds are 4.00, and (ii) in the second wager, the outcomes' odds are 1.60 and 2.10, and the backed outcome's odds are 1.60.

In Figure IA.2, we compute all possible prizes and corresponding probabilities of the lotteries associated with each of these three bets. The probabilities written in red color beside the branches of the lottery representations are the implied probabilities. For example, in the event with outcome odds 1.33 and 2.80, the implied probabilities are $\frac{1/1.33}{1/1.33+1/2.80} \approx 0.68$ and $\frac{1/2.80}{1/1.33+1/2.80} \approx 0.32$, respectively.

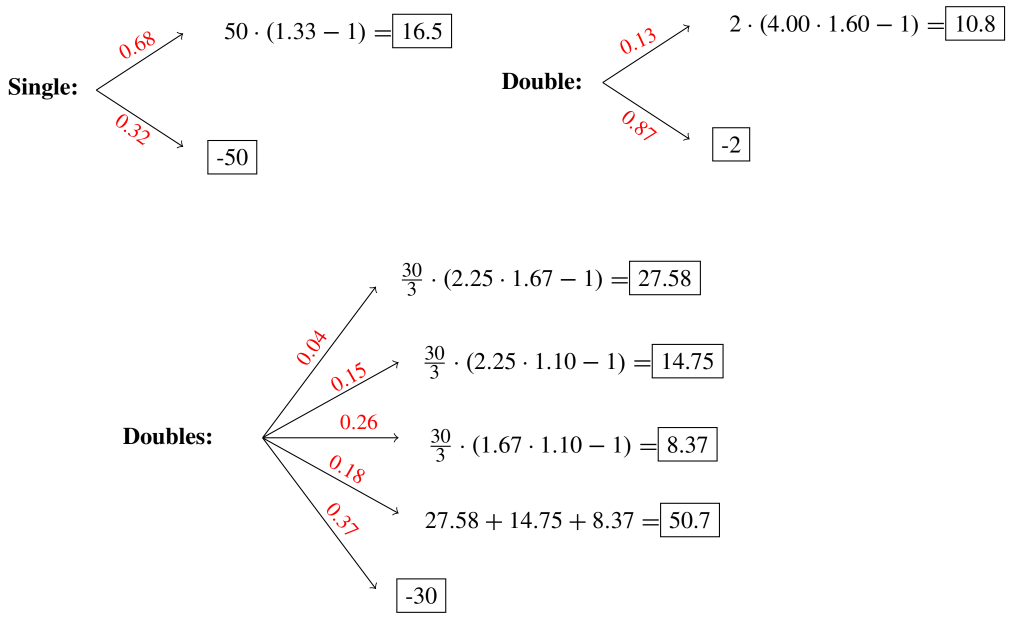


Figure IA.2: Lottery representations of various bets.

IA.3 Trace plots of MCMC draws

As explained in Section 1.3 of the paper, to estimate our model we derive the joint posterior distribution of the model parameters conditional on the data. Since this joint posterior cannot be calculated analytically, we obtain information about it by drawing from it using a Markov chain Monte Carlo (MCMC) algorithm.

Using this algorithm, we make 5 million draws from which we discard the first 10% as burn-in and retain every 50$^{\text{th}}$ after that to mitigate serial correlation. These draws form a Markov chain with stationary distribution equal to the joint posterior. In Figure IA.3, we present trace plots—i.e., plots of the retained draws against the iteration number—of the proportions of individuals that belong to each behavioral type. In Figure IA.4, we present trace plots of the population means and variances of the model parameters; for convenience of presentation, we present these trace plots for the case in which heterogeneity in the (transformed) parameters is simply modeled through a single normal distribution, rather than for the case in which individuals are allocated to four different utility types. Both these sets of plots indicate no problems with convergence.

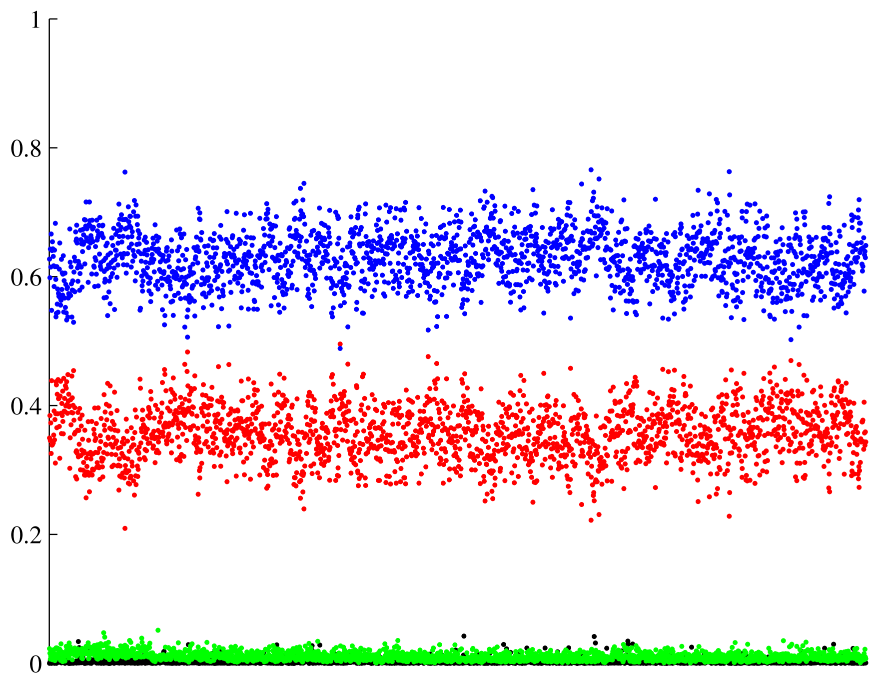


Figure IA.3: Trace plots of the MCMC draws for the population proportions of individuals belonging to (i) the type that exhibits no loss aversion and no probability weighting (at the bottom, using black dots), (ii) the type that exhibits no loss aversion (second from the top, using red dots), (iii) the type that exhibits no probability weighting (second from the bottom, using green dots), and (iv) the type that exhibits both loss aversion and probability weighting (at the top, using blue dots).

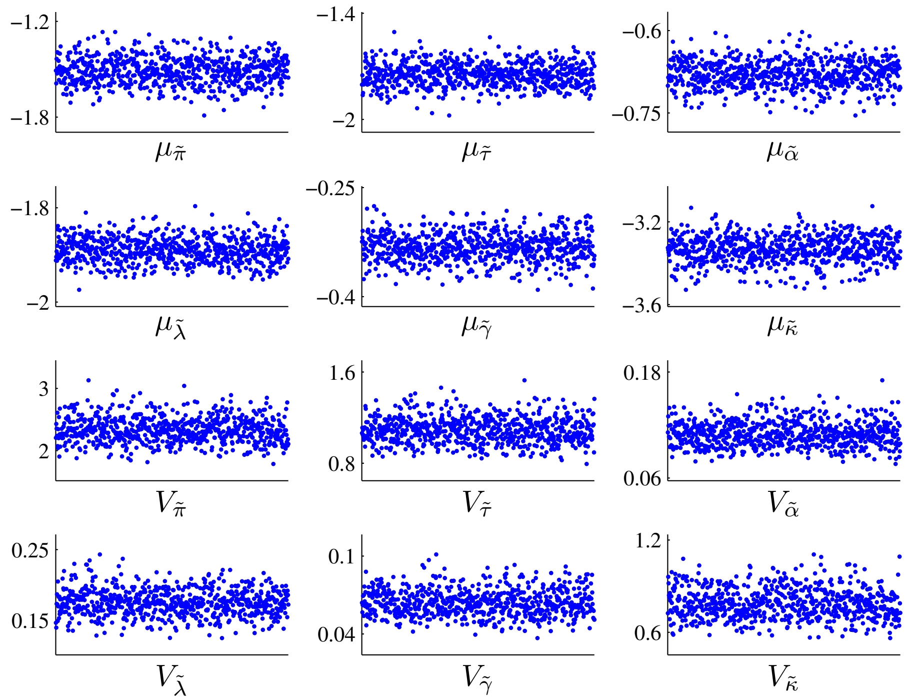


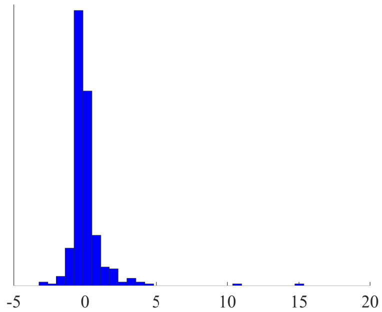
Figure IA.4: Trace plots of the MCMC draws for the population mean and variance of $\tilde{\pi}$, $\tilde{\tau}$, and $\tilde{\boldsymbol{\rho}}$.

IA.4 Model fit

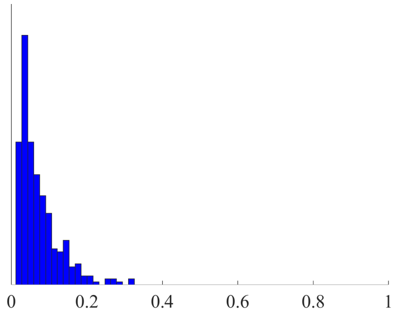
Here, we discuss how well our model fits the observed data. First, we show that the stochastic utility component needed to explain the observed choices is small. Normalizing to one the utility difference between the two degenerate lotteries with certain payoff of €0 and €1, respectively, we estimate the population median of the inverse scale τ of the random component to be 1.53. This is comparable to the value estimated in some lab experiments, and implies, e.g., that an individual making a choice between two lotteries whose utility differs by 1, chooses the less-preferred lottery with probability 0.19.[3] Even though the stochastic component of utility needed to fit the model to the data is small, it is important because it helps explain both why individuals make different risky choices at different times and why they sometimes place bets and sometimes do not. As we see in Figure IA.5a, for more than half of the individuals, their best chosen lottery has a slightly negative certainty equivalent, most often a few cents below zero. This means that these individuals are—according to the deterministic component of utility—almost indifferent between the safe and the chosen lottery, and that a very small stochastic component of utility (e.g., from utility of gambling) is sufficient to explain their choice to place a bet. Furthermore, as we discuss next, while the chosen risky lotteries are among the most desirable alternatives according to the deterministic component of utility, they are rarely *the* most desirable alternative.

Next, we assess model fit by examining if, for the preferences we estimate for each individual, the deterministic component of utility ranks his chosen lotteries among the most desirable alternatives. First, for each observed choice of each individual, we use his estimated preferences to calculate the chosen lottery's rank among the alternatives, and we divide this rank by the number of alternatives to produce a normalized rank that is close to 0 for desirable and close to 1 for undesirable lotteries. For each individual, we average this normalized rank across his choices, and in Figure IA.5b we plot the histogram of this mean ranking across individuals. For the median individual, the chosen lotteries are on average more preferred than 95% of the alternatives. Then in Figure IA.5c we plot a histogram, across all choices of all individuals, of the certainty equivalent (CE) difference between the chosen lottery and each of the alternatives (i.e., the individual-specific most-preferred lottery, second-most-preferred lottery, etc.). We see that the CE difference between the chosen and the most-preferred alternative is close to zero, meaning that, in utility terms, the chosen lottery is close to the most desirable alternative. Furthermore, the CE difference between the chosen and most alternative lotteries is quite high, indicating that the model does a good job of separating the desirable from the undesirable alternatives.

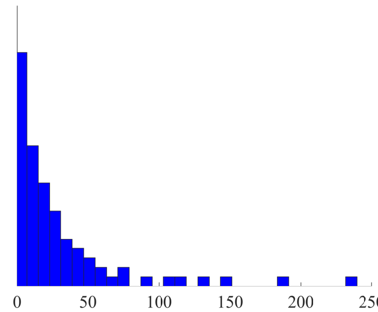
[3]For example, von Gaudecker, van Soest and Wengstrom (2011) find that an individual choosing between two lotteries whose utility differs by 1 would choose the less-preferred one with probability 0.45.



(a) Certainty equivalent difference, best chosen minus safe lottery.



(b) Mean ratio ranking of chosen lottery.



(c) Mean certainty equivalent difference, chosen minus alternatives.

Figure IA.5: Illustrations of goodness of fit of our model. Panel (a) plots the histogram, across individuals, of the certainty equivalent difference between the best chosen lottery and the safe lottery. Panel (b) plots the histogram, across individuals, of the mean ratio ranking of the chosen lotteries; for choices that are more (less) preferred, the ranking is close to 0 (1). Panel (c) plots the histogram, across alternatives, of the certainty equivalent difference between the chosen lotteries and the alternatives, averaged across choices and individuals.

Finally, we examine whether our model fits the data better than (i) a model in which lotteries are chosen randomly; (ii) a model in which all individuals have the Tversky and Kahneman (1992) parameters; and (iii) a model in which all individuals have the population mean parameters, so essentially the no-heterogeneity case. In Table IA.1, we present the average (across individuals) log likelihood from each model as well as the Deviance Information Criterion, which includes a penalty for model complexity. The results show that our model fits the data significantly better than all three alternatives. We also note that comparisons between models with only a subset of the CPT features are implicitly part of the estimation of our mixture model, which can be viewed as a tool for model selection analysis. As we show in Section 3 of the paper, the model with no loss aversion (the model with all CPT features) best explains observed behavior for one-third (two-thirds) of our individuals.

Table IA.1: Comparison of Our Model with Alternatives

	$\log L$		DIC
Model	Mean	Median	
Random Choice	-251.76	-150.51	503.52
Tversky-Kahneman Parameters	-227.85	-133.15	455.70
No Heterogeneity	-210.94	-121.19	421.88
Our Model	-179.27	-106.98	362.10

Mean and median log likelihood, across individuals, and Deviance Information Criterion (DIC) for our model and three alternatives: (i) one in which lotteries are chosen randomly; (ii) one in which all individuals have the Tversky and Kahneman (1992) parameters; and (iii) one in which all individuals have the population mean parameters. DIC $= \overline{D(\boldsymbol{\beta})} + p_D$, where $D(\boldsymbol{\beta}) = -2\log p(\boldsymbol{x}, \boldsymbol{y}\,|\boldsymbol{\beta})$, $\overline{D(\boldsymbol{\beta})} = \mathrm{E}[\,D(\boldsymbol{\beta})|\,\boldsymbol{x}, \boldsymbol{y}]$, and $p_D = \overline{D(\boldsymbol{\beta})} - D(\bar{\boldsymbol{\beta}})$, with $\boldsymbol{\beta}$ the model parameters and $\bar{\boldsymbol{\beta}} = \mathrm{E}[\boldsymbol{\beta}|\boldsymbol{x}, \boldsymbol{y}]$, i.e., the posterior mean.

IA.5 Application to portfolio choice/diversification

In this section, we provide more details on how we solve the portfolio choice problem described in Section 4 of the paper and we present some additional results.

In the asset-allocation problem we consider, an individual solves

$$\max_{\omega_S,\omega_F} U\left(r_p; \alpha, \lambda, \gamma, \kappa\right)$$
$$\text{s.t. } r_p = r_f + \omega_S r_S + \omega_F r_F,$$

where all quantities are defined in Section 4 of the paper. We solve this problem for each individual in our sample, using nominal annual returns for r_f, r_S, and r_F, and the posterior mean estimates of his preference parameters (α_n, λ_n, γ_n, and κ_n). To solve this problem, we construct a two-dimensional grid spaced at 1% increments over the range of values for the weights ω_S and ω_F, and we pick the weights that maximize the objective U for each individual. We approximate the objective by generating draws for the rates of return r_f, r_S, and r_F, and numerically calculating the integrals.

We generate 1 million draws for r_f, r_S, and r_F as follows. We obtain, for the period January 1975 to December 2011, data on the risk-free (the one-month Treasury Bill) rate and on stock returns from the Center for Research in Securities Prices (CRSP), and on U.S. equity mutual fund returns (net of fees, expenses, and transaction costs) from the CRSP Survivorship-Bias-Free U.S. Mutual Fund Database.[4] To generate each draw, we randomly pick a month t in the period January 1975 to January 2011, and then we randomly pick a stock among all the stocks and a fund among all the funds in our data in month t.[5] For the year starting in t, we calculate the rate of return for the risk-free asset and the excess rate of return for the chosen stock and fund. We reduce survivorship bias as follows: If a stock's/fund's return is missing from the data starting at some point during the year, we assume that with probability 0.5 it lost all its value hence its return for the year is -1, and with probability 0.5 its return information was removed from the data for a different reason hence we replace it with another stock/fund drawn randomly. The mean/median/standard deviation of our draws for r_f, r_S, r_F are 5.51%/5.23%/3.26%, 12.07%/4.69%/79.60%, and 11.49%/12.54%/24.52%, respectively. Using these draws, we can calculate the corresponding draws of the portfolio return r_p for each combination of weights ω_S and ω_F. Then, we can approximate the objective $U\left(r_p; \cdot\right)$ by calculating the weighted distribution function F_p on a linear grid of 10,000 values from slightly below the minimum to slightly above the maximum draw of r_p, calculating the corresponding numerical derivatives at these values, and finally calculating the numerical integral over the whole range.

To understand how our estimated distribution of preference parameters translates to the distribution of

[4]Since CRSP treats different share classes of the same fund as separate funds, we identify funds' share classes (using our own fund-name matching algorithm as well as the MF Links database from Wharton Research Data Services) and we compute each fund's monthly return as the weighted average of the returns of its component share classes, with weights equal to the beginning-of-month total net asset value of each class.

[5]The random draws of month t, the stock, and the fund are done with replacement. The probability of drawing each month is the same for all months, and the probability of drawing each stock (fund) equals the market capitalization (net asset value) of each stock (fund) divided by the total for all stocks (funds) active in month t. Alternatively, we could draw among all active stocks and funds with equal probability; our results using this alternative generating process are very similar to the ones we present here.

optimal portfolios we present in Figure 7, we analyze how each preference parameter affects the optimal direct and indirect investment in equity. In Figures IA.6 and IA.7, we present the optimal weights on the stock and fund as a function of the measure α of value-function curvature and of loss aversion λ, for various combinations of values for the measures of the curvature γ and elevation κ of the probability weighting function. We see that a higher α and a lower λ increase the total investment in equity and tilt the equity investment toward an undiversified portfolio and away from a well-diversified one. Furthermore, while the effect of γ and κ on total investment in equity is not monotonic, comparing Figure IA.7a with IA.7b and Figure IA.7c with IA.7d, we see that both a lower γ and a higher κ unambiguously tilt the equity investment toward an undiversified portfolio and away from a well-diversified one. Thus, the considerable heterogeneity we estimate in α and λ implies that half of the individuals have a combination of high enough α and/or low enough λ such that they optimally invest a significant fraction of their financial wealth in equity. Furthermore, the heterogeneity we estimate in all preference parameters accounts for the wide heterogeneity in the optimal split of individuals' equity investment between the stock and the fund.

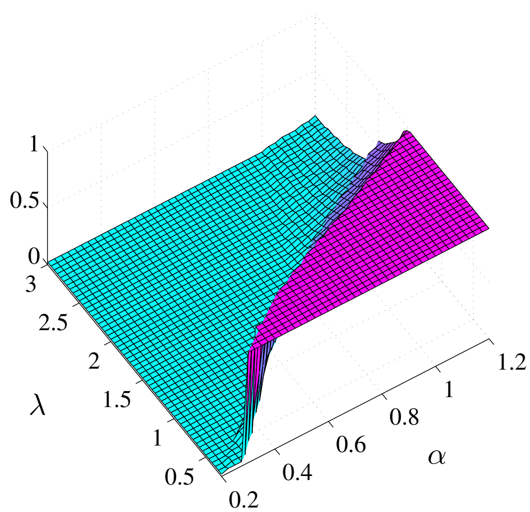
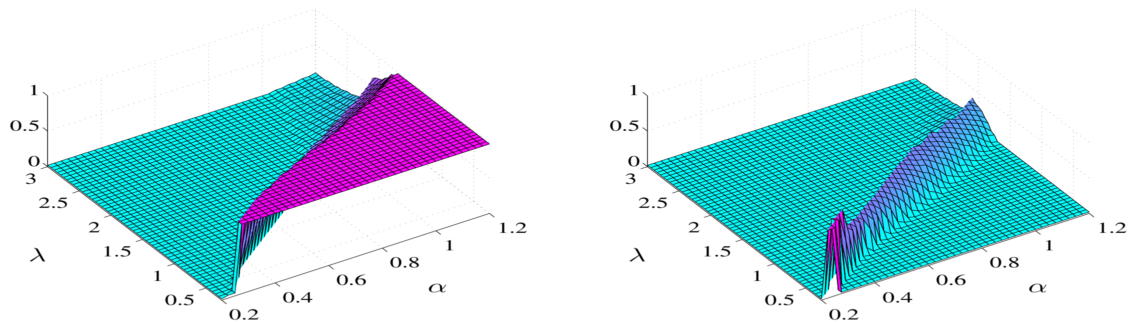
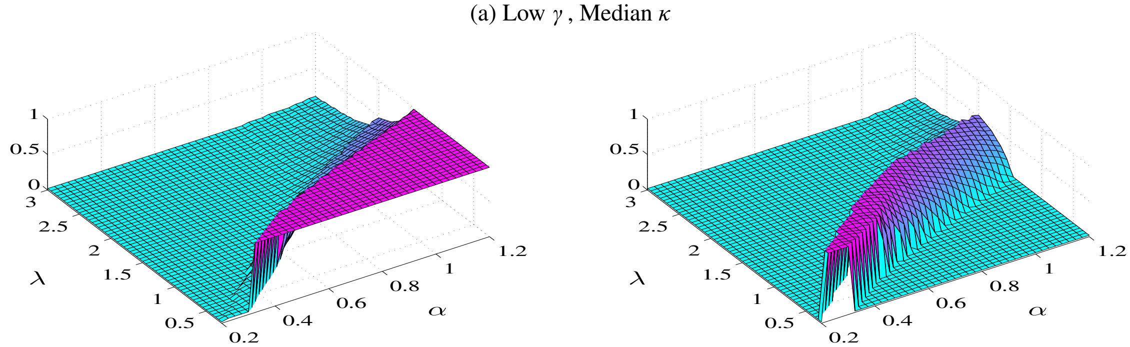


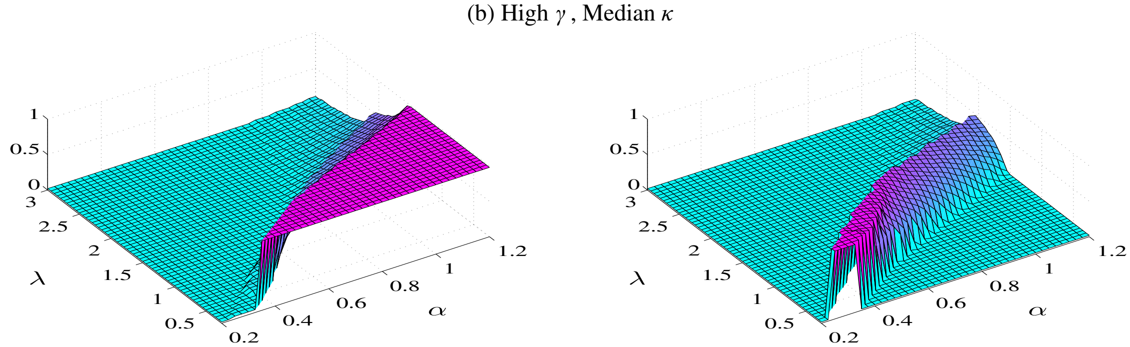
Figure IA.6: Optimal weights on the stock (in the left graph) and fund (in the right graph) as a function of α and λ, for the median values of γ (i.e., $\gamma = 0.84$) and κ (i.e., $\kappa = 2.14$).



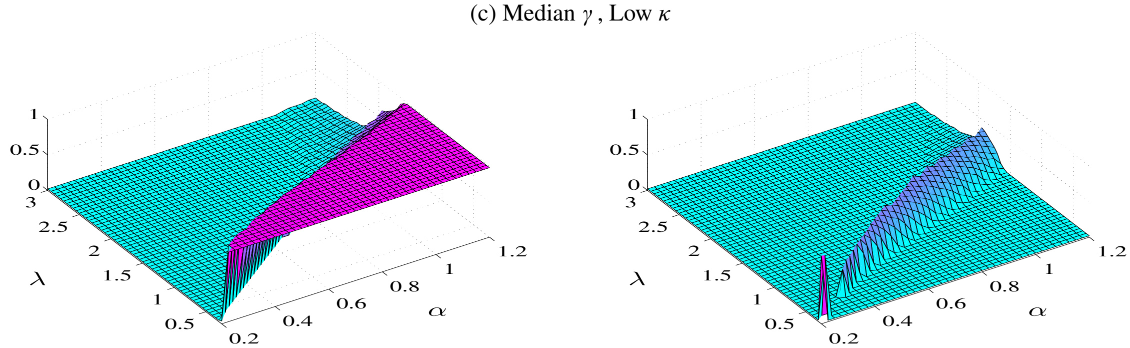
(a) Low γ , Median κ



(b) High γ , Median κ



(c) Median γ , Low κ



(d) Median γ , High κ

Figure IA.7: Optimal weights on the stock (in the left graph of each panel) and fund (in the right graph of each panel) as a function of α and λ, for various combinations of values of γ and κ across Panels (a) through (d). Panel (a) plots the optimal weights for $\gamma = 0.76$ (the 25$^{\text{th}}$ percentile of the estimated γ distribution) and $\kappa = 2.14$ (the median of the estimated κ distribution); Panel (b) for $\gamma = 0.92$ (the 75$^{\text{th}}$ percentile of the estimated γ distribution) and $\kappa = 2.14$; Panel (c) for $\gamma = 0.84$ (the median of the estimated γ distribution) and $\kappa = 1.61$ (the 25$^{\text{th}}$ percentile of the estimated κ distribution); and Panel (d) for $\gamma = 0.84$ and $\kappa = 3.38$ (the 75$^{\text{th}}$ percentile of the estimated κ distribution).

IA.6 Prior sensitivity analysis

In this section, we conduct a prior sensitivity analysis. We check the sensitivity of our results to varying the *marginal* priors of the population parameters, and also to varying the *prior predictive densities*, which compound all the marginals.

First, we present results from the sensitivity analysis in which we vary the scale $\underline{\boldsymbol{Z}}$ of the prior distribution for $\boldsymbol{V}^q_{\tilde{\rho}}$, from our baseline value ($\boldsymbol{I}$), to a low value ($0.1\boldsymbol{I}$), to a high value ($10\boldsymbol{I}$). In Figure IA.8, we plot the estimated population densities for each of these values for $\underline{\boldsymbol{Z}}$. We see that the prior has some effect on the estimated densities for the CPT parameters. However, the variances of all parameters are bounded far away from 0 under all prior specifications, suggesting that the heterogeneity we estimate is not driven by the choice of our priors. Moreover, while we vary the prior variances by an order of magnitude, the posterior variances vary by much less than that, indicating that the posteriors should be close to our baseline.

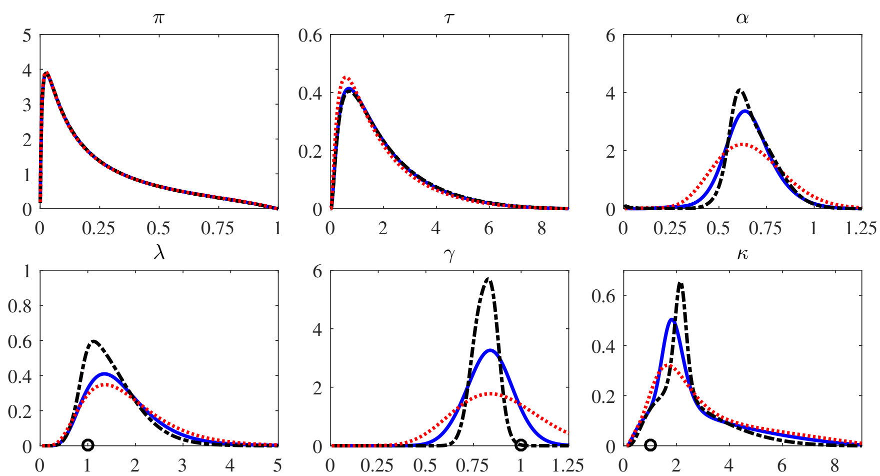


Figure IA.8: Prior sensitivity analysis, for different values of the scale $\underline{\boldsymbol{Z}}$ of the prior distribution for $\boldsymbol{V}^q_{\tilde{\rho}}$. We plot the estimated population densities from our baseline model, which has $\underline{\boldsymbol{Z}} = \boldsymbol{I}$ (in blue solid lines), from a model with small-variance priors, which has $\underline{\boldsymbol{Z}} = 0.1\boldsymbol{I}$ (in black dash-dotted lines), and from a model with large-variance priors, which has $\underline{\boldsymbol{Z}} = 10\boldsymbol{I}$ (in red dotted lines). The proportion with no loss aversion is 36% for the baseline priors, 27% for the small-variance priors, and 34% for the large-variance priors. The circles at $\lambda = 1$ and at $\gamma = 1$, $\kappa = 1$ represent point masses that correspond to the proportions of individuals with no loss aversion and with no probability weighting, respectively. The proportion with no probability weighting is 1% for the baseline priors, 7% for the small-variance priors, and 1% for the large-variance priors.

Second, we show that varying the priors for the utility type proportions $\boldsymbol{h}$ does not affect their posteriors, indicating that the priors do not drive our finding that a very substantial 36% of the individuals do not exhibit loss aversion. Specifically, we replace the symmetric $\mathcal{D}(\mathbf{1})$ prior with the asymmetric $\mathcal{D}(1, 1, 1, 13)$ prior, which underweights component probability vectors with higher values for the proportion of individuals with no loss aversion and overweights component probability vectors with higher values for the proportion of individuals who exhibit all features of prospect theory. In Figure IA.9, we present the estimated population densities of the model parameters that correspond to each of these two prior specifications.

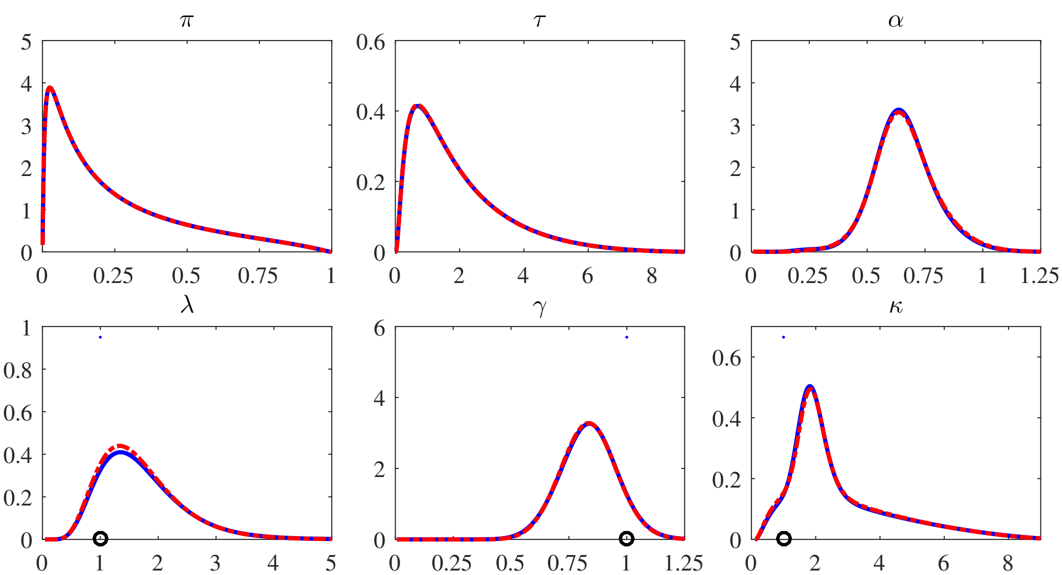
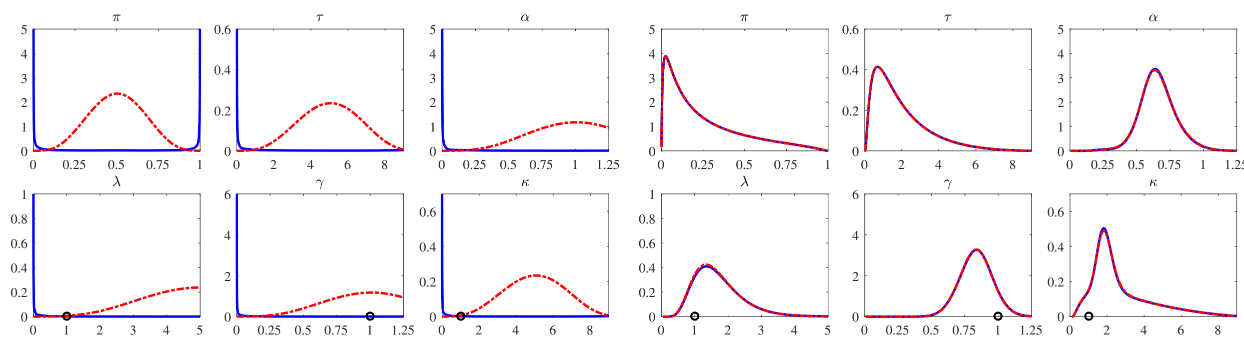


Figure IA.9: Estimated population densities of the model parameters, for the baseline prior specification that assumes a symmetric Dirichlet prior $\mathcal{D}(\mathbf{1})$ for the population proportions $\boldsymbol{h}$, and for an alternative prior specification that assumes an asymmetric Dirichlet prior $\mathcal{D}(1, 1, 1, 13)$. The estimated densities are plotted in solid blue for the baseline prior, and in dash-dotted red for the alternative prior. The circles at $\lambda = 1$ and at $\gamma = 1$, $\kappa = 1$ represent point masses that correspond to the proportions of individuals with no loss aversion and with no probability weighting, respectively. The proportion with no loss aversion is 36% for the baseline prior and 33% for the alternative prior, and the proportion with no probability weighting is 1% for both priors.

Finally, we check the sensitivity of the posterior predictive distributions of the preference parameters to their prior predictive distributions. We find that a drastic change in the prior predictive density—from the baseline that is flat almost everywhere, except for values close to the endpoints for bounded distributions, to an alternative that is concentrated over a smaller range of values away from the boundaries—has almost no effect on the posterior predictive density for all parameters, indicating that the data indeed contain information about the model parameters. To achieve the desired change in the prior predictive distribution, we vary some of the baseline hyperprior parameters for the population mean and variance of the preference parameters. Specifically, in the baseline priors we use parameters $\underline{\boldsymbol{\kappa}} = \mathbf{0}$, $\underline{\boldsymbol{K}} = 100\boldsymbol{I}$, $\underline{\zeta} = 6$, $\underline{\boldsymbol{Z}} = \boldsymbol{I}$, and in the alternative specification for which we present results here we use $\underline{\boldsymbol{K}} = 0.5\boldsymbol{I}$ (and $\underline{\boldsymbol{\kappa}} = \mathbf{0}$, $\underline{\zeta} = 6$, $\underline{\boldsymbol{Z}} = \boldsymbol{I}$, as in the baseline). The former parameter values (combined with our other prior parameters) imply that the prior predictive density is almost flat everywhere, except for values close to the endpoints for bounded distributions, while the latter imply that the prior predictive density is more concentrated over a small range of values away from the boundaries. As we show in Figure IA.10, the change in the prior predictive density is drastic but the effect on the posterior predictive density is very small.



(a) Prior predictive density, baseline vs. alternative.

(b) Posterior predictive density, baseline vs. alternative.

Figure IA.10: Plots of the prior and posterior predictive densities of π, τ, and $\boldsymbol{\rho}$ for the baseline prior specification ($\underline{\boldsymbol{\kappa}} = \mathbf{0}$, $\underline{\boldsymbol{K}} = 100\boldsymbol{I}$, $\underline{\zeta} = 6$, $\underline{\boldsymbol{Z}} = \boldsymbol{I}$) and for an alternative ($\underline{\boldsymbol{\kappa}} = \mathbf{0}$, $\underline{\boldsymbol{K}} = 0.5\boldsymbol{I}$, $\underline{\zeta} = 6$, $\underline{\boldsymbol{Z}} = \boldsymbol{I}$). In Panel (a), we plot the prior predictive densities for the baseline prior specification (in solid blue) and for the alternative prior specification (in dash-dotted red). The proportion with no loss aversion is 50% for both priors, and the proportion with no probability weighting is also 50% for both priors. In Panel (b), we plot the posterior predictive densities for the baseline prior specification (in solid blue) and for the alternative prior specification (in dash-dotted red). In both panels, the circles at $\lambda = 1$ and at $\gamma = 1$, $\kappa = 1$ represent point masses that correspond to the proportions of individuals with no loss aversion and with no probability weighting, respectively. The proportion with no loss aversion is 36% for both posteriors, and the proportion with no probability weighting is 1% for both posteriors.

References

von Gaudecker, H, A van Soest, and E Wengstrom. 2011. "Heterogeneity in risky choice behavior in a broad population." *American Economic Review*, 101(2): 664–694.

Tversky, A, and D Kahneman. 1992. "Advances in prospect theory: Cumulative representation of uncertainty." *Journal of Risk and Uncertainty*, 5(4): 297–323.