Jakusch, Sven Thorsten

Working Paper
On the applicability of maximum likelihood methods: From experimental to financial data

SAFE Working Paper, No. 148

Provided in Cooperation with:
Research Center SAFE - Sustainable Architecture for Finance in Europe,
Goethe University Frankfurt


This Version is available at:
http://hdl.handle.net/10419/146784

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes.
You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.
If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
Sven Thorsten Jakusch

On the Applicability of Maximum Likelihood Methods: From Experimental to Financial Data

SAFE Working Paper No. 148
Non-Technical Summary

Microeconomic modeling of investors’ financial decision making and its results crucially depend on the mathematical shape of the underlying preference function as well as its parameterization. The majority of such models rely on expected utility, first established in the seminal work of John von Neumann and Oskar Morgenstern in 1947. Due to the shortcomings of the expected utility theory in conciliating empirical evidence with theoretical predictions, researchers proposed alternative and generalized utility concepts, such as Rank-dependent, Prospect Theory and Cumulative Prospect Theory as a refinement of the former to improve descriptive accuracy of these models. This raises the question of which assumption about the proposed utility specification is valid to properly characterize risk preferences in financial markets. Solutions to the question of how to identify the best fitting utility model in controlled experimental environments have been proposed and conducted by e.g. John Hey and Christ Orme, however, to the best of our knowledge, no study had yet embarked on a comparable endeavor for behavior in financial markets.

This paper addresses whether and to what extent econometric methods used in experimental studies can be adapted and applied to financial data to detect the best-fitting preference model. To address the research question, we implement a frequently used nonlinear probit model and base our analysis on a tailor-made simulation study. In detail, we simulate trading sequences for a set of utility models and try to identify the underlying utility model and its parameterization used to generate these sequences by maximum likelihood.

We find that for a very broad classification of utility models, this method provides acceptable outcomes. Yet, a closer look at the preference parameters reveals several caveats that come along with typical issues attached to financial data and that partly seem to drive our results. In particular, deviations are attributable to effects stemming from multicollinearity and coherent under-identification problems, where some of these detrimental effects can be remedied up to a certain degree by adjusting the error term specification and thus the assumption of the likelihood function.

Furthermore, additional uncertainty stemming from changing market parameter estimates affects the precision of our estimates for risk preferences and cannot be simply remedied by using a higher standard deviation of the error term or a different assumption regarding its stochastic process. Particularly, if the variance of the error term becomes large, we detect a tendency to identify Prospect Theory as utility model providing the best fit to simulated trading sequences regardless of the utility function used to generate these trading sequences. We also find that a frequent issue, namely serial correlation of the residuals, does not seem to be significant.

However, we detect a tendency to prefer nesting models over nested utility models, which is particularly prevalent if Rank-dependent Utility and Exponential Power-Utility models are estimated along with CRRA utility models.
ON THE APPLICABILITY OF MAXIMUM LIKELIHOOD METHODS: FROM EXPERIMENTAL TO FINANCIAL DATA

SVEN THORSTEN JAKUSCH

Abstract. This paper addresses whether and to what extent econometric methods used in experimental studies can be adapted and applied to financial data to detect the best-fitting preference model. To address the research question, we implement a frequently used nonlinear probit model in the style of Hey and Orme (1994) and base our analysis on a simulation study. In detail, we simulate trading sequences for a set of utility models and try to identify the underlying utility model and its parameterization used to generate these sequences by maximum likelihood. We find that for a very broad classification of utility models, this method provides acceptable outcomes. Yet, a closer look at the preference parameters reveals several caveats that come along with typical issues attached to financial data, and that some of these issues seems to drive our results. In particular, deviations are attributable to effects stemming from multicollinearity and coherent under-identification problems, where some of these detrimental effects can be captured up to a certain degree by adjusting the error term specification. Furthermore, additional uncertainty stemming from changing market parameter estimates affects the precision of our estimates for risk preferences and cannot be simply remedied by using a higher standard deviation of the error term or a different assumption regarding its stochastic process. Particularly, if the variance of the error term becomes large, we detect a tendency to identify SPT as utility model providing the best fit to simulated trading sequences. We also find that a frequent issue, namely serial correlation of the residuals, does not seem to be significant. However, we detected a tendency to prefer nesting models over nested utility models, which is particularly prevalent if RDU and EXPO utility models are estimated along with EUT and CRRA utility models.

Date: 24 December, 2013.

Key words and phrases. Utility Functions, Model Selection, Parameter Elicitation.

House of Finance, Goethe University Frankfurt, Grueenbergplatz 1, D-60323 Frankfurt am Main, Germany. Sven Jakusch is a doctoral student at House of Finance, Goethe University Frankfurt and Senior Quantitative Consultant at Ernst & Young GmbH. We are grateful for comments by Andreas Hackethal, Glenn Harrison, John Hey, Steffen Meyer and Chris Orme. We gratefully acknowledge research support from the Research Center SAFE, funded by the state of Hessen initiative for research LOEWE. The corresponding authors can be reached by sven-jakusch@yahoo.de. Please note that parts of this paper were written when Sven Jakusch was working at Ernst & Young Wirtschaftsprüfungsgesellschaft GmbH, however, any views, statements or opinions expressed in this paper are solely those of the authors and not related to Ernst & Young.
1. Introduction

Microeconomic modeling of investor financial decision making and its results crucially depend on the mathematical shape of the underlying preference function as well as its parameterization. The majority of such models rely on expected utility, first established in the seminal von Neumann and Morgenstern (1947). Due to the shortcomings of the expected utility theory in conciliating empirical evidence with theoretical predictions, researchers proposed alternative and generalized utility concepts, such as Rank-dependent Utility (Quiggin (1982)), Prospect Theory as conceptualized by Kahneman and Tversky (1979) and Cumulative Prospect Theory as a refinement of the former (posed in Tversky and Kahneman (1992)) to improve descriptive accuracy of these models. This raises the question, which assumption about the proposed utility specification is valid to properly characterize risk preferences in financial markets? Solutions how to identify the best fitting utility model in controlled experimental environments have been proposed and conducted by e.g. Hey and Orme (1994) and Laury and Holt (2005), however, to the best of our knowledge, no study had yet embarked on a comparable endeavor for behavior in financial markets. Financial data, in contrast to experimental data, have certain merits but also come with significant disadvantages that might require some modifications of the methods adopted from experimental economics. In particular, such data comprise revealed rather than stated preferences (Train (2009)), offer a considerable sample size but might suffer from unobservable factors beyond the control of the researcher, first-order autocorrelation in the time series as well as between unobservable factors, under-identification problems regarding the utility models and a certain degree of multicollinearity (Campbell et al. (1997)) and the introduction of additional uncertainty stemming from the way investors tend to extrapolate past returns into the future (Andreassen and Kraus (1990)) and carry over accumulated gains and losses over time. We assess the compatibility of both the econometric concepts adopted from the experimental literature, namely customized maximum likelihood methods, based on modeling an additional error term on top of the individuals decision rule as introduced by Hey and Orme (1994), and the effects of selected features of financial data, particularly the trading behavior from individual investors. Therein, we identify and analyze potential problems that arise when using trade data such as multicollinearity, the effects of additional uncertainty regarding the stochastic properties of the likelihood function, autocorrelation and the identifiability of the true but unknown functional shape of an investor’s utility function. Those problems mentioned can be generated by the way investors obtain estimates to approximate uncertain financial outcomes (Kahneman and Tversky (1973), Andreassen and Kraus (1990)) and by carrying forward accrued returns over time.

The paper is organized as follows. We outline the research question and provide an overview of the current relevant experimental literature to frame the topic in the second chapter. In the third chapter, we sketch a frequently applied likelihood approach for utility model identification and present the inherent statistical properties that can be expected to hold in experimental data. This allows us to identify and highlight several weak points in the widely applied likelihood approach with regard to financial data and to show from which factors potential problems may arise. In chapter four we focus predominantly on the problems we identified in a preliminary likelihood analysis of the results from the application of the maximum likelihood method from a simulation study before we present and discuss each of these problems in detail, particularly how these results from a model selection procedure can be affected. In this chapter, we analyze in more detail the factors...
suspected to alter the surface of the likelihood function, to yield unreliable estimators for risk preference parameters, and to cause breakdowns in certain numerical search algorithms, thus affecting model selection results. We elaborate to what extent the above-mentioned effects arise and alter the results of the utility model identification strategy. Our findings confirm that the additional uncertainty that yields to a modification of the stochastics of the likelihood function, introduced by the unknown stochastic process stemming from carrying forward intermediate gains and losses, multicollinearity and under-identification interferes with the precision of our estimators and thus the identification of the underlying utility model. Autocorrelation, on the other hand, appears to predominantly affect selection of the error term specification. However, if the error term specification interacts with model selection in a way that, by coincidence, captures part of the effects of the additional uncertainty and of multicollinearity, estimation of the variance of the error term, an additional nuisance parameter, can supplement and enhance the identification of the correct utility model specification.

2. Preferences in financial markets and experiments

Empirical and theoretical research in finance has placed great emphasis on the mathematical specification of the utility function of an individual decision maker, usually concluding that most individuals are indeed risk averse, despite the fact that notable exceptions exist.\(^1\) Findings from experimental and empirical studies provide a multifaceted picture in this matter, as they are based on various methods and data sets, and thus they are difficult to compare directly and yield results that are virtually impossible to conciliate. Despite an apparent consensus about the general relevance of risk aversion in the theory of financial decision making, the exact characterization of an investor’s utility function is a highly disputed topic. For example, regarding the classical Expected Utility paradigm, studies such as Friend and Blume (1975), Blume and Friend (1975), Schlarbaum et al. (1975), Morin and Suarez (1983) and Landskroner (1988) find evidence for decreasing absolute risk aversion (DARA) and constant relative risk aversion (CRRA) and estimate the coefficient for relative risk aversion to be approximately 2 or higher; this is incompatible with the assertions of logarithmic utility (Latane (1959), Hakansson (1971) and Markowitz (1976)), a conjecture supported by recent empirical studies based on household survey data.\(^2\) In contrast to this strand of the literature, experimental evidence outside the field of finance is vast and finds somewhat lower values, with ambiguous results. For example, Gordon et al. (1972), Kroll et al. (1988) and Levy (1994), who analyze the experimentees decision making in a portfolio choice context, show evidence for DARA and moderate support for either increasing relative risk aversion (IRRA) or CRRA, although other authors object that IRRA might be an artifact of the inherent positive-or-zero gain feature of such experiments (Levy (1994)) and that absolute risk aversion cannot be unambiguously recovered from actual choices (Wolf and Pohlman (1983)).

Although it appears that DARA and CRRA utility dominates financial markets, classical utility theories have been questioned due to their incompatibility with\(^1\)Friedman and Savage (1948), Markowitz (1952), Tversky and Kahneman (1992) and Kahneman and Tversky (1979).

\(^2\)Although Brunnermeier and Nagel (2008) conclude that logarithmic utility may provide an appropriate description of financial market risk aversion, Guiso and Paiella (2008) and Chiappori and Paiella (2011) detect signs of DARA and CRRA with (highly dispersed) coefficients of relative risk aversion above 2 for a significant proportion of households.
Empirical phenomena, such as the observed equity premium\(^3\), and on the observed violation of their inherent axiomatic properties (Allais (1953), see also Edwards (1996), Barberis and Thaler (2003), Glaser et al. (2004), Shefrin (2008), Wang (2006), Broihanne et al. (2008), the recent works by Wakker (2010) and Barberis (2013)). To address the latter issue, early experimental studies such as Preston and Baratta (1948), Edwards (1953), and Edwards (1954) reveal that decision makers systematically violate the independence axiom of von Neumann and Morgenstern (1947), thus concluding that subjects decide in discord with physical probabilities and seem to apply decision weights to making choices. These findings prompted development of generalized expected utility theories, such as Dual Theory (Yaari (1987)) and creation of Rank-dependent Utility (RDU) as advocated by Edwards (1962), Karmarkar (1978), Karmarkar (1979), Quiggin (1982) and Wakker (1994), whereas another strand of the literature proposed modifications of the utility function itself (e.g., Friedman and Savage (1948), Markowitz (1952), Kahneman and Tversky (1979), Wakker and Tversky (1993)). According to these generalized expected utility models, risk aversion is now not only determined by the curvature of the utility function, but also dependent on the shape of the decision weight attached to the alternatives of the choice set. Despite the fact that increasing marginal utility causes a degree of discomfort for economists (Yaari (1965)), recent empirical and experimental evidence provides further support for these alternative utility models (e.g., Hakansson (1970), Hershey and Schoemaker (1980), Tversky and Kahneman (1991), Tversky and Kahneman (1992), Rabin (2000), Rabin and Thaler (2001), Levy and Post (2005), Wakker (2010)), although these concepts are not beyond criticism (Levy and Levy (2002b), see also Wakker (2003)).

Empirical studies that address whether alternative utility theories, particularly both versions of Prospect Theory, in which risk aversion is now captured by three different parameters, are effective and present in financial markets predominantly focus on selected features, such as the effects of reference points and loss aversion on financial decision making - features mostly argumentatively connected to individual investor trading behavior. For illustrational purpose, the pioneering Shefrin and Statman (1985) establishes a possible link between Prospect Theory and observed financial decisions, namely the Disposition Effect, by basing its reasoning on these aforementioned characteristics, although this link has been questioned recently (e.g., Barberis and Xiong (2009), Linmainaa (2010), Vlcek and Hens (2011) and Barberis (2012)).\(^4\) Other studies focus on the interdependence between Prospect Theory and option exercise behavior (Heath et al. (1999), Potheshman and Serbin (2003)), the behavior of futures traders in real markets (Locke and Mann (2000), Locke and Mann (2005) and Coval and Shumway (2005)) and experiments

\(^3\)For a critique on the empirical findings for risk aversion based on their inconsistency with the equity premium, see Mehra and Prescott (1985), Mankiw and Zeldes (1991), Benartzi and Thaler (1995), Blake (1996), Cocherlakota (1996), Goetzman and Ibbotson (2005) and Mehra (2008)

\(^4\)If individual preferences follow the predictions of Prospect Theory, phenomena such as the Disposition Effect should be observable in other environments. In fact, evidence for the Disposition Effect has been found among individual investors in the stock market (Scharbaum et al. (1978), Ferris et al. (1988), Odean (1998), Weber and Camerer (1998), Odean (1998), Odean (1999), Garvey and Murphy (2004), Jordan and Dilitz (2004), Lehenkari and Perttunen (2004), Frazzini (2006), Dhar and Zhu (2006)) and other environments, such as in the financial advice of stock brokers (Shapira and Venezia (2001)), the behavior of futures traders (Heisler (1994), Frino et al. (2004), Coval and Shumway (2005) as well as Locke and Mann (2005)), IPO trading volume (Kantua (2004)), real estate markets (Genesove and Mayer (2001)), insurance contracts (i.e.Schoemaker and Kunreuther (1979), Camerer and Kunreuther (1989)), and observed risk behavior in laboratory environments for stocks (Weber and Camerer (1998), Oehler et al. (2003), Lee et al. (2008)) and monetary endowments (see Chui (2001)).
(Haigh and List (2005), see also Harrison and Rutstrom (2009), who argue that these effects may also be consistent with CARA under variable risk aversion), as well as observed behavior in real estate markets (e.g. Genesove and Mayer (2001)).

In light of such numerous, partly contradictory empirical evidence regarding the mathematical nature of individual investor preferences, comparison of utility functions is an ongoing topic in the experimental literature, as documented by Lattimore et al. (1992), Hey and Orme (1994) and Abdellaoui (2000). A major breakthrough in the quest to find underlying preferences is Hey and Orme (1994), who assessed various parametric utility functions at the level of individual subjects using specific customized maximum likelihood procedures (Orme (1995)).

Maximum likelihood methods in general capture the idea that individuals err in their decision making, a presumption that is not explicitly recognized in previous methods on preference estimation (e.g., as in McCord and DeNeufville (1986), Currim and Sarin (1989) and others). These maximum likelihood methods have been enhanced and widely applied, generating a large number of studies addressing the best-fitting utility model, predominantly finding evidence favoring prospect theory models. However, these results were criticized for their artificial setting concerning the payoff-structure (Laury and Holt (2005)), as well as the way in which relevant information was presented (Kahneman and Tversky (1973)). Naturally, the question arises as to whether these well-established methods can be adapted and applied to financial data, assuming the decision process and data structure are similar to that obtained in the laboratory. In the next section, we present the frequently applied maximum likelihood approach for utility model selection in detail and elaborate on the econometric peculiarities that accompany financial data.

3. AN ECONOMETRIC MODEL OF FINANCIAL DECISION MAKING

Characterizing risk preferences of individual investors in financial markets typically involves extensive individual-level analysis, if one refrains from applying mixed models to capture the heterogeneity in preferences in the manner considered to be effective to unravel the respective utility functions and their parameterizations (e.g., Hey and Orme (1994), Laury and Holt (2005)). For this reason, the emphasis of empirical and experimental studies, particularly in the context of finance and asset markets, has shifted away from aggregated views that describe investors

---

5In contrast to the abundance of experimental evidence, studies directly addressing the best-fitting utility function in financial markets are surprisingly scarce. A notable exception is Blackburn and Ukhov (2006), who applies a modified approach from Jackwerth (2000) at the level of individual stocks to draw conclusions on the underlying utility function from the direction of the sign of the estimated pricing kernels. Blackburn and Ukhov find evidence for utility functions similar to those proposed by Friedman and Savage (1948), Markowitz (1952) and Kahneman and Tversky (1979), where classical utility functions such as CRRA prevail only in 3 out of 41 cases.

6For instance, Blondel (2002) fits linear, power, and exponential forms of utility to experimental data. This author finds strong evidence in favor of the power and the exponential function. In the same vein, Stott (2006) also find a best fit for power and exponential functions, while quadratic and linear specifications perform poorly. A further comparison between the first two functional forms shows, in line with Blondel (2002), that power specifications fit even better to the experimental data. Results from further experiments, in which maximum likelihood methods are applied, are mostly consistent with an inverse S-shaped probability weighting function (Wu and Gonzalez (1996), Wu and Gonzalez (1999), Abdellaoui (2000), Bleichrodt and Pinto (2000), Abdellaoui et al. (2005)), moreover, consistent with a concave value function in the domain of gains, corresponding to prospect theory, which is also backed in recent studies that address the best-fitting functional form (see, e.g., Wakker (2008)). The properties of diminishing sensitivity toward variations in areas of gains are further confirmed in Wakker and Denneffe (1996), Fox and Tversky (1998) and Fennema and van Assen (1999), whereas evidence of risk-seeking in the realm of losses is already shown by Fishburn and Kochenberger (1979).
as a unified whole, usually modeled as a representative investor (Duffie (2001), Back (2012) or Munk (2013)), to disaggregated models in which the analysis is conducted on an individual level (for an early reference see Hensher and Johnson (1981)). In comparison to aggregate data, information on individual decisions is usually characterized by infrequent and discrete observations, by greater variation in each factor and less covariation among these factors, due to aggregation procedures, by summing the individual observations (Train (1986)). A variety of econometric methods have been developed to address apparent discreteness, such as discrete choice models (Amemiya (1975), McFadden (1980), Amemiya (1981), Amemiya (1985)), which have been popularized by Ben-Akiva and Lerman (1985), Train (1986), Train (2009) and advanced for estimation of non-linear arguments, such as utility functions in a customized maximum likelihood model (Harrison and Rutstrom (2008) and de Palma et al. (2008)) as proposed by Hey and Orme (1994).

To test the effectiveness of such a likelihood-based model selection procedure, it is necessary to identify common and distinctive features in the design of laboratory and financial market environments. Experimental studies on utility functions, which apply such discrete choice models for preference measurement by customizing an underlying likelihood function, usually are similar in multiple aspects regarding underlying assumptions on the decision process. Narrow framing is one central assumption, as it allows to define a finite and exhaustive set of alternatives via two mutually exclusive options, satisfying the requirements for a discrete choice set (Train (1986), Train (2009)) such that discrete choice models are applicable. To define the set of choice options, it is commonly assumed that, for each investment decision of an individual investor, the respective choice set is spanned by a risky asset and an investment in an assumingly riskless money market account in place of a representative riskless asset, yielding a gross return of $R_{f,t}$. Given the usual lottery-type design, the price of the risky asset, essentially any stock traded by the investor over the respective period, is assumed to be subject to a stochastic binomial process (Cox et al. (1979) and Rendleman and Bartter (1979), also Hull and White (1988)) in which two disjunct states $S$ of the world can be identified, yielding a gross return of $R_{S,t}$.

---

7 This is important, because the precision of estimation generally increases with sample size and variance of the variables entering the model, and decreases with its covariance. Further, standard econometric tools, such as regression analysis, implicitly assume a set of continuous variables—an assumption appropriate for aggregated models but that seems to be problematic when underlying factors on an individual investor level are the focus of interest.

8 Empirical studies in finance also indicate that investors allocate different streams of income, such as dividends and cash flows resulting from corporate actions and other stocks (Shefrin and Statman (1984), Baker and Wurgler (2004)), to different mental accounts (Thaler (1985)). Further, the tendency to evaluate risky lotteries separately, known as narrow framing (Barberis and Huang (2001), Barberis et al. (2001), Berkelaar et al. (2004), Gomes (2005), Barberis and Huang (2009)) is in line with Shefrin and Statman (1985), complementing recent studies on individual investors that examine trading decisions for each stock separately (see Odean (1998), Odean (1999), Barber and Odean (2000), Barberis and Huang (2001), Grinblatt and Keloharju (2001a), Grinblatt and Keloharju (2001b), Barber and Odean (2002), Dhar and Kumar (2002), Hong and Kumar (2002), Zhu (2002), Grinblatt and Han (2005), Liu (2006), Frazzini (2006), Barber and Odean (2008)).

9 In the upside state $U$, associated with some unknown physical probability $p_t > 0$, indicated by an index $t$ for the time, the stock price follows a rise and yields an upside return $R_{U,t} > 1$, whereas in the downside state $D$ with corresponding probability $1 - p_t$, the stock declines, generating a downside return $0 \leq R_{D,t} < 1$. 

6
The expected payoffs and accrued returns of the risky asset constitute a central distinctive feature between financial markets, in which parameterization of the underlying return distribution is unknown and intermediate (paper) gains and losses are followed up, and experiments, in which gains and losses are not carried forward to the next lottery task to avoid strategic hedging behavior and where payoffs of the lotteries are clearly presented to the experimentee. There is ample evidence that investors in financial markets experience difficulty in recognizing and learning the true but unknown market parameterizations, especially if they vary over time (Ehm et al. (2012)). To approximate financial payoffs, investors form their expectations on the outcomes of the risky asset by applying several mental shortcuts. DeBondt (1993) mentions that investors may consider recent past returns to be representative to formulate their expectations about the future to approximate financial payoffs (Kahneman and Tversky (1973), for evidence from stock markets see Andreassen (1987), Andreassen (1988) and Andreassen and Kraus (1990)). Considering that this mental pattern transforms the payoffs $R_{S,t}$ into expected profits extrapolated from the past over some lookback period, it generates an implicit correlation of these values as time proceeds and beliefs on market parameters are updated in each $t$.\footnote{In doing so, we distinguish from the representativeness bias, whereupon investors base their judgments on stereotypes and seek out patterns in returns or prices (Weber and Camerer (1998), Shefrin (2008)). The intuition here is that, due to extrapolation bias with short horizons, investors may buy stocks whose price has recently increased, especially when following a myopic trading strategy, which contradicts mean reversion expectation (Zuchel (2001)). This is backed in empirical studies, for example Grinblatt and Keloharju (2000) and Kaustia (2010). These authors find that Finnish investors bought past winners and sold past losers, thus revealing a trend-following trading strategy, which is not consistent with an expectation of mean-reverting stock prices (see also Kaniel et al. (2008)). Dhar and Kumar (2002) investigate the price trends of stocks bought by more than 62,000 households using discount brokerages, and conclude that investors prefer to buy stocks that have recently enjoyed an abnormal return.}

Studies on empirical dynamic programming suggest that individual investors face computational difficulties in determining the optimal trading strategy (see Eckstein and Wolpin (1989), Rust (1994) and Adda and Cooper (2003) for surveys). According to these studies, investor behavior is more likely reconcilable with a discrete decision process (Rust (1992)), which is found to be reflected in the stock market (Thaler et al. (1997) and Gneezy and Potters (1997), see also Normandin and St-Amour (2008)).\footnote{Further, under generalized utility concepts, dynamic programming with nonlinear decision weights can generate suboptimal and time-inconsistent results, as shown by Machina (1989) for non-expected utility in general, Nielsen and Jaffray (2004) for Rank-dependent Utility RDU and Barberis (2012) and Ebert and Strack (2012) for Cumulative Prospect Theory CPT, respectively.} To formalize the decision process, the utility an individual investor obtains from a money market account is denoted as $V_k(W_t,R_{f,t}|θ_k)$ for utility model $k$, whereas the utility resulting from the risky asset is symbolized by the expression $V_k(W_t,R_{S,t}|θ_k)$ with $S \in \{U; D\}$. This allows the introduction of a parameter set $θ_k$ to represent the utility-specific parameters of utility model-type $k$, which in turn, due the discretionary (myopic) decision process, leads to the behavioral assumption that the investor invests a positive amount in the risky asset if

$$V_k(W_t,R_{S,t}|θ_k) ≥ V_k(W_t,R_{f,t}|θ_k) \quad (3.1)$$

holds. Discrete choice models as in Train (1986), Rust (1994) and Train (2009) are constructed around the assumption that only a minority of attributes that drive purchase and selling decisions in financial markets are observable. Consequently, the utility function has the additively separable decomposition

$$V_k(W_t,R_{S,t}|θ_k) = U_k(W_t,R_{S,t}|θ_k) + ε, \quad (3.2)$$
which, combined with (3.1), implies that the investor holds the stock if the difference in utilities, abbreviated as $\Delta_k(U_k| \theta_k)$, is positive. There are two main purposes of the stochastic component $\epsilon$: first, the error term should fully capture hidden factors that affect the observed variations in the attractiveness of the respective stock without the necessity to explicitly model other (potentially unobservable) variables or data imperfections (Cramer (1986), also Rust (1994)). Secondly, without adding the error term $\epsilon$ to $\Delta_k(U_k| \theta_k)$, the lack of error in behavior will yield imprecise estimates for $\theta_k$.

To illustrate the importance of the error term regarding the second argument, consider the case of an investor, whose decision is exclusively based on $\Delta_k(U_k| \hat{\theta}_k)$. While for each set of $\theta_k$ there exists an unique optimal decision for the investor, the converse is not true. For any set of optimal decisions reflected in the trading sequence of this investor there exists a set of parameters consistent with those decisions. In terms of the likelihood function, if no error term is added to $\Delta_k(U_k| \hat{\theta}_k)$, then for any parameter set $\theta_k$ the observations are either consistent with the utility model or not. If they are not consistent with the utility model under consideration, the likelihood is zero and $\log L(\Delta_k(U_k| \hat{\theta}_k))$ is (theoretically) $-\infty$. However, if the observations are in accord with the utility model, then the likelihood is one and $\log L(\Delta_k(U_k| \hat{\theta}_k))$ is zero. As a consequence, $\log L(\Delta_k(U_k| \hat{\theta}_k))$ is oscillation between zero and $-\infty$. Therefore, modeling and estimation of an individual’s trading behavior should contain an additional element $\epsilon$. Note that, based on the predictability of $R_{t,t}$ and the fact that the utility of the risk-free investment carries no uncertainty per se, $\epsilon$ stemming from the riskfree asset is assumed to be zero.\textsuperscript{12}

To obtain the maximum likelihood function, conditional choice probabilities are derived given the stochastic properties of the error term, which is frequently assumed to be normally distributed (Hey and Orme (1994) and Carbone and Hey (2000)) as $\epsilon \sim N(0, \sigma^2)$ with density according to $\phi(\epsilon) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}(\epsilon/\sigma)^2}$. The derivation of the conditional choice probabilities requires defining an index $I_{k,t} := P[\Delta_k(U_k| \theta_k) + \epsilon \geq 0]$, taking the value 1 if the condition in the brackets is met and zero otherwise.\textsuperscript{13} The resulting choice probabilities are usually denoted as $\Phi(\Delta_k(U_k| \theta_k)/\sigma)$, where by the finiteness of the choice set, the probability of investing in the riskless asset is $1 - \Phi(\Delta_k(U_k| \theta_k)/\sigma)$ (or $-\Phi(\Delta_k(U_k| \theta_k)/\sigma)$). In this case, $\Phi$ denotes the cumulative normal density function, but it may be substituted by any other distribution, such as the lognormal (as in Booij et al. (2009)) or logistic distribution (Harrison and Rutstrum (2008), Train (2009)). The term $\sigma^2_k$ denotes the (heteroscedastic) variance of the error term on a daily basis, estimated as a nuisance parameter (Pawitan (2001)) along with $\theta_k$ to absorb potential effects resulting from the lagged structure of the estimated market parameters.

\textsuperscript{12}This is a minor technicality, as it avoids the necessity to evaluate all elements of the covariance matrix of errors (see Train (2009)). Further note that by assuming hedonic framing, the covariance in errors between the stocks traded in a portfolio can be ignored.

\textsuperscript{13}The probability (Rust (1994)) of buying or holding the risky asset is thus given as

$$p(\Delta_k(U_k| \theta_k) \geq 0) = p[I[\Delta_k(U_k| \theta_k) + \epsilon \geq 0] = 1]$$

$$= \int_{-\infty}^{\infty} I[\Delta_k(U_k| \theta_k) + \epsilon \geq 0] \phi(\epsilon) d\epsilon$$

$$= \int_{-\infty}^{\infty} \Phi(\Delta_k(U_k| \theta_k)/\sigma) \phi(\epsilon) d\epsilon,$$

satisfying the conditions if $\Delta_k(U_k| \theta_k) \to \infty$, the choice probability converges to unity and approaches zero if $\Delta_k(U_k| \theta_k) \to -\infty$. 8
within $\Delta_k(U_k(\theta_k))$ (Dhrymes (1971), Cramer (1986)) and of its error terms. Combining the normal distribution of the error term with the binary choice feature of the discrete choice setting leads to a likelihood function $\log L(\Delta_k(U_k(\theta_k)))$ similar to a non-linear in arguments probit model (see Thurstone (1927), case V, for an early reference from the field of psychometrics and Marschak (1960) for a transition in terms of utility). Experimental and empirical studies, which commonly apply likelihood methods to identify utility functions, implicitly use the convenient properties of maximum likelihood estimators (such as the frequently cited Jullien and Salanie (2000)), whereupon maximizing $\log L(\Delta_k(U_k(\theta_k)))$ for each of the $K_k$ elements of $\theta_k$ and $\sigma$ provides estimators $\hat{\theta}_{k|n,t}$ for a given sample size, indicated by the number of stocks $n$ and trading days $t$, which are consistent, asymptotically efficient, and moreover, asymptotically normally distributed.

To identify the best fitting underlying utility function of type $k$, insights from likelihood theory provide the key for the selection of the utility model that explains observed data best. According to Fisher (1922) and Kullback (1968), the maximized likelihood function $\log L(\Delta_k(U_k(\hat{\theta}_k)))$ contains information for each utility model $k$ on the relative fit of this model to observed data. To distinguish the $k$ utility models from one another, classical likelihood theory suggests that the utility model with the highest maximized likelihood $\log L(\Delta_k(U_k(\hat{\theta}_k)))$ fits observed data best (Kullback (1968), Akaike (1973), Schwarz (1978), Amemiya (1980), Pawitan (2001) and Burnham and Anderson (2004)). However, note that $\log L(\Delta_k(U_k(\hat{\theta}_k)))$

---

14 In addition, the economic intention of the heteroscedasticity feature of the error term $\epsilon$ allows us to disintegrate factors with varying impact, dependent on whether the decision at hand is a purchase or a sale, thus agreeing with Odean (1999), Glaser and Weber (2007) and Statman et al. (2006), whereupon purchase decisions may be motivated by factors other than sell decisions. For example, overconfident investors may suffer from biased beliefs about the anticipated returns they expect to generate by trading stocks even if these investors had average performance in the past (Odean (1999), Barber and Odean (1999) and Glaser and Weber (2007)), as such investors are induced to buy stocks more readily.

15 The overall likelihood function of an investor of utility type $k$ can accordingly be expressed as $\log L(\Delta_k(U_k(\theta_k))) = \sum_{t \in T} \sum_{I_k,t} I_{k,t} \log p_{k,t}(\Delta_k(U_k(\theta_k)))$ in which $p_{k,t}(\Delta_k(U_k(\theta_k)))$ denotes the respective conditional probabilities as defined above. Given the binary choice assumption, the log-likelihood function can be explicitly written as $\log L(\Delta_k(U_k(\theta_k))) = \sum_{t \in T} \log \left( \frac{\phi \left( \frac{\Delta_k(U_k(\theta_k))}{\sigma} \right) }{\phi \left( \frac{-\Delta_k(U_k(\theta_k))}{\sigma} \right) } \right)^{I_{k,t}} \left( \frac{1}{\phi \left( \frac{\Delta_k(U_k(\theta_k))}{\sigma} \right) } \right)^{1-I_{k,t}}$ in which we omit constant combinatorial terms since they add no further information about $\theta_k$. An alternative expression, which explicitly recognizes the dichotomy of $I_{k,t}$, is obtained if the log-likelihood function is decomposed as $\log L(\Delta_k(U_k(\theta_k))) = \sum_{t} \log \phi \left( \frac{\Delta_k(U_k(\theta_k))}{\sigma} \right) + \sum_{t} \log \phi \left( \frac{-\Delta_k(U_k(\theta_k))}{\sigma} \right)$, which immediately carries over to the notation of the score vector and the Hessian matrix. Expressing the log-likelihood function this way is helpful in organizing the computations but cumbersome if one wishes to derive the information matrix such that the notation used in this paper appears to be more convenient although it seems ostentatious at first sight.

16 In the appendix, we provide further details concerning these properties of the maximum likelihood methodology in the context of utility models for the interested reader. Concerning these estimators, given that certain regularity conditions of $\log L(\Delta_k(U_k(\theta_k)))$ hold, Lehmann (1983) elaborates that $\hat{\theta}_{k|n,t}$ has certain convenient properties, that is, that its estimators cannot be located on a boundary without violating the regularity of the likelihood function. Further, the information matrix is bounded and positive and thus satisfies the characteristics of a variance measure given that a second-order Taylor expansion of $\log L(\Delta_k(U_k(\theta_k)))$ is sufficient and valid. Experiment design can ensure that the presupposition, according to which the single likelihood functions are indeed independent (e.g., Post et al. (2008) and others), is maintained and consequently the asymptotic features of $\hat{\theta}_{k|n,t}$ are preserved. For instance, current studies on preferences in game shows allow for carrying forward gains and losses (Post et al. (2008)) in a way that preserves the likelihood properties but may require simulation methods to establish the empirical distribution function upon which the likelihood approach is constructed.
doesn’t necessarily have to be exactly zero for the perfect fitting model: Consider the case where the true model is \( y \sim N(0, 1) \) and we fit the model \( y \sim N(\mu, \sigma) \) by estimating \( \mu \) and \( \sigma \) by maximum likelihood. Furthermore assume data are generated by sampling from a standard normal distribution using simulation. Even if maximizing of the likelihood identifies zero and one as parameter estimates, the log \( \log L(\Delta_t(U_k|\theta_k)) \) of each observation is not going to be zero. It is rather going to be the standard normal density evaluated at \( y \), so the total log \( \log L(\Delta_t(U_k|\theta_k)) \) will not be zero in general.\(^{17}\)

Ranking all utility models according to \( \log L(\Delta_t(U_k|\theta_k)) \) and choosing the model with the highest likelihood as model selection criterion is usually not recommended for utility model selection as the maximized likelihood function is subject to overfitting, tendentially favoring multiparameter utility models such as Prospect Theory (3 parameters) to Expected Utility (one parameter) (Carbone and Hey (1994), Hey and Orme (1994), Carbone and Hey (1995) and Stott (2006)). Instead, literature on model selection suggests sorting utility models according to the Akaike Information Criterion (AIC) that controls explicitly for varying number of parameters instead of using the maximized \( \log L(\Delta_t(U_k|\theta_k)) \) (Akaike (1973), Akaike (1974), Bozdogan (2000), Pawitan (2001) and Burnham and Anderson (2004)). The AIC is commonly expressed as

\[
AIC = \frac{2 \log L(\Delta_t(U_k|\hat{\theta}_k))}{nt} + \frac{2K_k}{nt},
\]

according to Akaike (1974) in the representation of Amemiya (1980), where dividing by \( nt \), the number of observations in terms of trading days \( t \) and traded stocks \( n \), corrects for the different number of observations and where \( K_k \) denotes the rank of \( \theta_k \), representing the number of parameters to be estimated in utility model \( k \). Due to the general finiteness of our dataset, we apply the corrected Akaike Information Criterion (AICC), defined by

\[
AICC = \frac{2 \log L(\Delta_t(U_k|\hat{\theta}_k))}{nt} + \frac{2K_k}{nt} + \frac{2K_k(K_k + 1)}{nt(nt - K_k - 1)},
\]

as first proposed by Sugiuira (1978) for Ordinary-Least-Square (OLS) regressions (for a discussion of the original version of the Akaike Information Criterion AIC and AICC as model selection criterion we refer to Burnham and Anderson (2004)), which replaces the penalty term of AIC by its exact term for bias adjustment, resulting in a greater penalty for models with additional parameters in comparison to the original AIC. We provide a more formal outline in the Appendix.

4. A Simulation Study of Likelihood based Utility Model Selection

Thus far, we sketched the elements of the likelihood theory as applied in experimental studies addressing individual preferences and argued that applying maximum likelihood estimation provides asymptotically efficient and unbiased estimators. Moreover, making use of the characteristics of the likelihood function \( \log(L(\Delta_t(U_k|\theta_k))) \), particularly the structure of its surface - its elevation and its steepness, allows filtering for the best-fitting model. As pointed out, experimental

\(^{17}\)Note that even if we remove randomness from our data (i.e. the variation of the error term is zero), the likelihood of each of these identical observations will be the normal density evaluated at zero, which is not zero but a positive number (approx. 0.399). Thus, the overall logarithm likelihood is \( n t \cdot \ln(0.399) > 0 \) although we specified the correct model. Note that in our case, the observations are not equal to their means due to \( \Delta_t(U_k|\theta_k) \). I’m am grateful to John Hey who made me aware of this.
literature often assumes that each single likelihood function of \( \log(L(\Delta_t(U_k|\theta_k))) \) is independent. Given the tendency to extrapolate past moments into the future and to keep track of paper gains and losses, the interlacing of the various functionals of the likelihood function introduces an implicit dependency within and across \( \log(L(\Delta_t(U_k|\theta_k))) \), resulting in deficiencies in the surface of the likelihood function. In this case, the extend and direction of these effects are unclear and requires an investigation, how these adverse effects transmit to estimators for \( \hat{\theta}_{k|n,t} \) and how much these imprecisions can negatively affect the identification of the underlying utility mode.

With the intention to investigate the reliability of likelihood-based model selection procedures given financial data and the identification of factors that are detrimental for this purpose, we conduct a conceptually simple simulation study that consist of four steps: First, to control effects stemming from the financial times series, we simulate a series of prices and their returns with known stochastic characteristics and market parameters. In the second step, we change the perspective and take the position of an individual investor with known utility functions and risk parameters, who faces this set of hypothetical stocks, represented by the time series of returns. We estimate the unknown market parameters needed for the third step, in which the investor decides whether she wants to buy, hold or sell the hypothetical stocks according to inequality (3.1). From this step, we obtain a set of trading sequences (one for each hypothetical stock), that we use in the final step to identify the underlying utility function and infer its parameterization used to generate the trading sequence.

**STEP 1: Simulation of a Set of Time Series for given Market Parameters:**

To begin with, we simulate realizations of a sequence of returns based on a pre-specified stochastic Markovian process to avoiding inherent autocorrelation of our time series upfront. In total, we generated a set of 100 hypothetical stocks by simulating a series of identically and independently distributed (logarithmic) returns spanning 312 days each, denoted as

\[
\left\{ \left( R_{t,t+1} | \mu_\Delta(t), \Phi^{-1}(0,\sigma)\sqrt{\Delta(t)} \right) : t = 1, 2, \ldots, 312 \right\}.^{18}
\]

The time series of returns is modeled as discrete time Geometric Brownian motion by the inverse of the cumulative standard normal distribution \( \Phi^{-1}(0,\sigma) \) plus a trend \( \mu \), where the time step is set to \( \Delta(t) = 1 \). To calibrate the trend variable \( \mu \) and standard deviation \( \sigma \), we are guided by the results of Dimson et al. (2000), Dimson et al. (2003) as well as publicly available data for the German stock market and set the daily mean \( \mu = 1.0004 \) and volatility \( \sigma = 0.0247.^{19} \) With regard to the riskfree rate, we opt for a fixed daily net return of 0.0001, corresponding to a return of 0.67% assuming annual compounding. For the inverse of the cumulative standard normal distribution of each simulated return sequence, we defined the seed values of the random number generator in Stata Version 10.1 (Cameron and

---

18 Due to the rolling-window procedure of Step 2, where the unknown market parameters are estimated, we extended the time series of each stock by 60 days such that (technically) the length is 312; the final time series after Step 2 thus spans 252 days, which is a common approximation for the number of trading days in one year.

19 Gross returns and risk premia (see Fama and French (1993)) are obtained from [http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html#International](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html#International) as accessed on 02.12.2012 to obtain values for \( \mu \), which we use in the second step to match \( \sigma \) with historical gross appreciation rates to make simulated returns correspond more closely to their empirical equivalents.
Experimental econometrics for finance - Analysis of a Likelihood Approach

Trivedi (2005), Baum (2006)) by a Halton sequence (see Halton (1960)) based on the prime 11 to make sure the full spectrum of random numbers for the stochastic process is covered. As the draws from a Halton sequence tend to be negatively correlated with previous draws, this feature is beneficial to reduce simulation error, consequently less simulations are required to obtain statistically reliable results (Train (2009)). Exemplarily, Bath (2001) found that a sequence of 100 Halton draws provides more precise results for a mixed logit estimation than 1000 random draws. Table (1) illustrates the simulations conducted in Step 1.

Figure 1. Simulated Time Series of Prices

The figure below illustrates the development of simulated returns for 100 draws used in this simulation study. Market parameters are fixed at $\mu = 1.0004$ and $\sigma = 0.0247$ respectively. For this table, prices are calculated on a daily basis using the simulated returns generated in Step 1. The starting price of each stock was set to 100.

STEP 2: Estimation of Market Parameters from Simulated Time Series:

In the second step, we take the position of an individual investor and estimate the market parameters needed to fill the respective utility functionals. For this purpose, estimators for the mean $\hat{\mu}_t$ and volatility $\hat{\sigma}_t$ were obtained using a rolling-window procedure with a lookback horizon of $l = 60$ days. Values of upside and downside returns $\hat{R}_{D,t}$ and $\hat{R}_{U,t}$ were then derived from $\hat{\mu}_t$ and $\hat{\sigma}_t$ for each $t$, where

$A Halton sequence based on prime $h$ is defined as $s_{t+1} = \{s_t, s_t + \frac{1}{h}, s_t + \frac{2}{h}, \ldots, s_t + \frac{h-1}{h}\}$, where $s_{t+1}$ denotes the sequence at iteration $t+1$ of length $h$. The application of a Halton sequence constitutes a well-defined draw spanning the standard uniform density, as it systematically fills in the unit interval.

---

20A Halton sequence based on prime $h$ is defined as $s_{t+1} = \{s_t, s_t + \frac{1}{h}, s_t + \frac{2}{h}, \ldots, s_t + \frac{h-1}{h}\}$, where $s_{t+1}$ denotes the sequence at iteration $t+1$ of length $h$. The application of a Halton sequence constitutes a well-defined draw spanning the standard uniform density, as it systematically fills in the unit interval.
of the utility functionals are the commonly applied CRRA
following settings
Rank-dependent Utility (RDU)
roundtrip sequences for an investor of trading behavior with those generated by generalized expected utility and generate EXPO
R
W
STEP 3: Calculation of Trading Sequences for each Utility Function:
In the third step, given a vector \( \theta_k \) of predefined risk preference parameters for the respective utility function of type \( k \) of the simulated investors, we generate a variety of artificial trading histories in terms of roundtrip sequences (Shapira and Venezia (2001)), denoted accordingly as
\[
\{ (\hat{R}_{U,t}, \hat{R}_{D,t}, R_{F,t}, \hat{p}_t) : I_t \in [0; 1] \ \forall t = 1, 2, \ldots, 252 \}.
\]
The mathematical details of the utility functions used in this simulation study are relegated to the Appendix.\textsuperscript{21} To calculate accrued returns, we set the gross realized return to unity if the index is zero and add the realized logarithmic returns to \( W_t \) if the index is 1 until the last day before it switches back to zero, which we take as realized return being invested in the riskfree asset.\textsuperscript{21}

To substantiate the set of utility functions that specifies \( \Delta_t(U_k|\theta_k) \), we generate the trading patterns for an investor characterized as a Expected Utility Theory (EUT)-type investor of the CRRA form (A.3) with the following settings:
\[
\{ (U_{EUT}(W_t, R_{U,t}, R_{D,t}, R_{F,t}, p_t|\theta_{EUT} \in \{ \delta \})) \},
\]
where \( \theta_{EUT} : \{ \delta \in \{2\} \} \),

where we deliberately set \( W_t \) to zero at \( t = 1 \) and equal to the accrued return \( \hat{R}_{S,t} \) from the risky investment for \( t \in \{2, 3, \ldots, T\} \) for simplicity. Specifications of the utility functionals are the commonly applied CRRA-form and exponential power utility (EXPO) as proposed in Saha (1993). We contrast this simulated trading behavior with those generated by generalized expected utility and generate roundtrip sequences for an investor of Rank-dependent Utility (RDU)-type with the following settings
\[
\{ (U_{RDU}(W_t, R_{U,t}, R_{D,t}, R_{F,t}, p_t|\theta_{RDU} \in \{ \delta, \gamma \} )) \},
\]
where \( \theta_{RDU} : \{ \delta \in \{2\}; |\gamma \in \{0.65\} \} \)

\textsuperscript{21}This is a standard procedure (Ingersoll (1987)) and is widely applied as similar expressions can be found in Johnson et al. (1997), Barberis and Xiong (2009), Ebert and Strack (2009) and Johnson et al. (2012).
\textsuperscript{22}Note that this behavior can imply rational behavior, since \( \hat{p}_t \) also serves as maximum-likelihood estimator for the underlying true but unobservable probability \( p \) given a binomial distribution \( p_t = (t+1)p(t+1)(1-p)^t \). The solution for the estimator \( \hat{p} = \frac{21}{22} \) is derived by taking the logarithm and differentiating with respect to \( p \).
\textsuperscript{23}In detail, to generate a sequence of trades that resembles the lottery task features as it is the common framing in experimental studies, we define an index \( I(\Delta_t(U_k|\theta_k) + \epsilon \geq 0) \) similar to the index used to obtain the choice probabilities in the previous chapter, yielding a series of zeros and ones dependent on whether the investor holds the stock on a particular day. These sequences allow us to define so-called roundtrips, defined as rows of ones similar to the definition of Shapira and Venezia (2001), since the various trade-inventory rules Odean (1998) coincide for binary choice situations. The effects of different accounting principles on our results is outside of the scope of our paper.
\textsuperscript{24}According to our definition, the roundtrip sequence comprises only the row of ones under the fiction that a new mental account is opened if the investor sells and buys at a later point in time. This avoids dilution effects due to accruing riskfree rates in the estimation and having different starting values for \( W_t \), with the result that risk preferences would be measured at different locations of the utility functional, which we suspect to yield different estimates if \( W_t \) is high.
Table 1. Utility Models and Parameters $\theta_k$ used for Trading Sequences

The table provides a description of parameter set $\theta_k$, its settings used for generating the trading sequences and the key reference. Expected utility models are denoted as $EUT$, Rank-dependent Utility is denoted as $RDU$. For Simple Prospect Theory (Tversky and Kahneman (1992)) we use the denotation $SPT$, whereas Cumulative Prospect Theory (Tversky and Kahneman (1992)) is denoted as $CPT$. Furthermore, for $EUT$ and $RDU$ we use the denotation $CRRA$ for utility functionals with constant relative risk aversion and $EXPO$ to denote exponential power utility functions (Saha (1993)). For $SPT$ and $CPT$, we use the denotation $POWR$ to indicate models with kinked power-functionals and $CRRA$ if a power functional used to specify the prospect value functional (Gomes (2005)), where in addition $DHG0$ denotes value functionals as defined in DeGiorgi and Hens (2006).

<table>
<thead>
<tr>
<th>Set $\theta_k$</th>
<th>Interpretation</th>
<th>Key Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EUT</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CRRA</strong></td>
<td>$\delta = 2$</td>
<td>Risk Aversion</td>
</tr>
<tr>
<td><strong>EXPO</strong></td>
<td>$\delta = 2$</td>
<td>Risk Aversion</td>
</tr>
<tr>
<td>$\rho = 1$</td>
<td>Scaling Parameter</td>
<td>Saha et al. (1994)</td>
</tr>
<tr>
<td><strong>RDU</strong></td>
<td>$\delta = 2$</td>
<td>Risk Aversion</td>
</tr>
<tr>
<td>$\gamma = 0.65$</td>
<td>Weighting Parameter</td>
<td>Quiggin (1982), Tversky and Kahneman (1992)</td>
</tr>
<tr>
<td><strong>SPT</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CRRA</strong></td>
<td>$\alpha = 0.88$</td>
<td>Risk Sensitivity</td>
</tr>
<tr>
<td>$\lambda = -2.25$</td>
<td>Loss Aversion</td>
<td>Tversky and Kahneman (1991)</td>
</tr>
<tr>
<td>$\gamma = 0.65$</td>
<td>Weighting Parameter</td>
<td>Quiggin (1982), Tversky and Kahneman (1992)</td>
</tr>
<tr>
<td><strong>POWR</strong></td>
<td>$\alpha = 0.88$</td>
<td>Risk Sensitivity</td>
</tr>
<tr>
<td>$\lambda = -2.25$</td>
<td>Loss Aversion</td>
<td>Tversky and Kahneman (1991)</td>
</tr>
<tr>
<td>$\gamma = 0.65$</td>
<td>Weighting Parameter</td>
<td>Quiggin (1982), Tversky and Kahneman (1992)</td>
</tr>
<tr>
<td><strong>DGH0</strong></td>
<td>$\alpha^+ = 0.20$</td>
<td>Scaling Parameter</td>
</tr>
<tr>
<td>$\lambda^+ = 6.52$</td>
<td>Scaling Parameter</td>
<td>DeGiorgi and Hens (2006)</td>
</tr>
<tr>
<td>$\lambda^- = 14.7$</td>
<td>Scaling Parameter</td>
<td>DeGiorgi and Hens (2006)</td>
</tr>
<tr>
<td>$\gamma = 0.65$</td>
<td>Weighting Parameter</td>
<td>Quiggin (1982), Tversky and Kahneman (1992)</td>
</tr>
<tr>
<td><strong>CPT</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CRRA</strong></td>
<td>$\alpha = 0.88$</td>
<td>Risk Sensitivity</td>
</tr>
<tr>
<td>$\lambda = -2.25$</td>
<td>Loss Aversion</td>
<td>Tversky and Kahneman (1991)</td>
</tr>
<tr>
<td>$\gamma = 0.65$</td>
<td>Weighting Parameter</td>
<td>Quiggin (1982), Tversky and Kahneman (1992)</td>
</tr>
<tr>
<td><strong>POWR</strong></td>
<td>$\alpha = 0.88$</td>
<td>Risk Sensitivity</td>
</tr>
<tr>
<td>$\lambda = -2.25$</td>
<td>Loss Aversion</td>
<td>Tversky and Kahneman (1991)</td>
</tr>
<tr>
<td>$\gamma = 0.65$</td>
<td>Weighting Parameter</td>
<td>Quiggin (1982), Tversky and Kahneman (1992)</td>
</tr>
<tr>
<td><strong>DGH0</strong></td>
<td>$\alpha^+ = 0.20$</td>
<td>Scaling Parameter</td>
</tr>
<tr>
<td>$\lambda^+ = 6.52$</td>
<td>Scaling Parameter</td>
<td>DeGiorgi and Hens (2006)</td>
</tr>
<tr>
<td>$\lambda^- = 14.7$</td>
<td>Scaling Parameter</td>
<td>DeGiorgi and Hens (2006)</td>
</tr>
<tr>
<td>$\gamma = 0.65$</td>
<td>Weighting Parameter</td>
<td>Quiggin (1982), Tversky and Kahneman (1992)</td>
</tr>
</tbody>
</table>

with decision-weightings according to Quiggin (1982) ($QU82$) and Tversky and Kahneman (1992) ($KT92$), regarding the utility functional, we use $CRRA$ and $EXPO$. To test the sensitivity of the likelihood-based model selection with respect to non-standard utility functions, we also consider non-expected utility investors of $Simple Prospect Theory (SPT)$-type (Kahneman and Tversky (1979)) of the form

$$\{U_{SPT}(W_t, W_{RP}, \theta_{UL}, \theta_{DL}, \theta_{f, t}, \theta_{p}; \theta_{SPT} \in \{\alpha, \gamma, \lambda\}\},$$

where $\theta_{SPT} : \{\alpha \in [0.88]; \gamma \in [0.65]; \lambda \in [-2.25]\}$,

where we consider $W_{RP}$ to be located at the purchase price of the risky asset without inherent dynamics.

We accompany the results from the generalized expected utility models by the trading sequences of a $Cumulative Prospect Theory (CPT)$-type investor with the following settings

$$\{U_{CPT}(W_t, W_{RP}, \theta_{UL}, \theta_{DL}, \theta_{f, t}, \theta_{p}; \theta_{CPT} \in \{\alpha, \gamma, \lambda\}\},$$

where
which covers approximately the upper and lower bounds of estimated parameterizations from experimental studies. As for the \textit{SPT} case, we use a decision weight according to Quiggin (1982) and Tversky and Kahneman (1992) for the \textit{CPT} function. For \textit{SPT} and \textit{CPT}, we specified the functional form of the value functional according to a power function (Kahneman and Tversky (1979)) and contrast the results with an \textit{CRRA} functional (as in Barberis et al. (2001) and Gomes (2005)), exponential power functional (\textit{EXPO}) and a piecewise negative exponential power functional (\textit{DGH0}) as defined in DeGiorgi and Hens (2006). Table (1) provides an overview of the utility models used and their parameter settings in this simulation study.

As \textit{SPT} and \textit{CPT} type-investors are expected to be sensitive with respect to the prospect horizon, such that we calculate the respective utilities according to a predetermined forecast periods, denoted as $\tau$. Due to the fact that \textit{CPT} and \textit{SPT} coincide under $\tau = 1$ and are thus indistinguishable, we set $\tau = 20$. To avoid side relations among the elements of $\theta_k$ and to model decision errors as e.g. Carbone and Hey (1994) and Carbone and Hey (1995), we add a normally distributed error term $\epsilon$ to the difference of the respective utility of the risky stock and the riskless money market account, for which we deliberately set the standard deviation of the error equal to 0.01. On each day, the hypothetical investor invests in the risky stock whenever $\Delta_t(U_k|\theta_k) + \epsilon \geq 0$. If this condition is satisfied, we set the indicator $I[\Delta_t(U_k|\theta_k) + \epsilon \geq 0]$ to one and zero otherwise. The repetition of this for all 252 trading days of each time series of Step 1 yields the required trading sequences for the maximum likelihood estimation in Step 4.

**STEP 4: Estimation and Selection of Utility Function from Trading Sequences:**

In the final step, we take the trading sequences from Step 3, evaluate the likelihood function $\log(L(\Delta_t(U_k|\theta_k)))$ for each utility type $k$ and perform the model selection. In detail, for each of the 100 trading sequences of each investor of utility type $k$, we loop through all conceivable utility functions, for which we estimate the associated risk preference vector $\theta_k$ and the standard deviation of the error term $\sigma_\epsilon$. Regarding the latter, we transform $\sigma_\epsilon$ in the likelihood estimation by an exponential function (as described by Rabe-Hersketh and Everitt (2004), Chapter 13) to ensure that the ascertained estimator is strictly positive. We recover the estimator for $\sigma_\epsilon$ and the associated standard errors using the \texttt{nlcom} command in Stata version 10.1 (for details on the maximum likelihood estimation see Gould et al. (2006)). For each trading sequence of an investor of utility type $k$, every time the likelihood function has been evaluated for each conceivable utility function, we rank the utility models by sorting the corrected Akaike-criterion ($AICC$) before proceeding with the next trading sequence for this investor.

For linear-in-parameter logit models, McFadden (1974) shows that there exists a unique and global maximum of the likelihood function, however, due to utility models such as \textit{RDU}, \textit{SPT}, and \textit{CPT} and therefore the nonlinear structure of $\Delta_t(U_k|\theta_k)$, it cannot be expected that the likelihood function we face is well-behaved and can likewise be characterized by a unique global maximum. This has several consequences for the results of a model selection procedure: If it cannot be ruled out that $\hat{\theta}_{k|n,t}$ is the result of a stopped numerical search due to a local maximum in the likelihood function or a sufficiently flat region of $L(\Delta_t(U_k|\theta_k))$ (e.g., McCullough and Vinod (2003)), then estimates of $\theta_k$ are potentially located...
far from the true values. If nesting models are tested against nested ones, the imprecision in the estimation of \( \hat{\theta}_{k|m,t} \) thus may favor the former, since the respective parameter constraints cannot be ascertained properly. We address potential problems in the numerical search algorithm stemming from deficiencies in the surface of the likelihood function \( \log L(\Delta_t(U_k|\theta_k)) \) in two ways: Firstly, in accordance with suggestions from literature (Judge et al. (1985), there Appendix B, Ruud (2000) and Gould et al. (2006)) we modify the numerical search algorithm every five steps. Thus, for the numerical search algorithm required to evaluate \( \log(L(\Delta_t(U_k|\theta_k))) \), we run a Newton-Ralphson procedure for the first five steps. If no solution is obtained or the algorithm fails to converge, we switch to the Davidon-Fletcher-Powell algorithm (Fletcher (1980)) for the next five iterations to push the estimates outside the critical section of the likelihood function and then return to the former technique.\(^{25}\) With regard to the number of iteration steps, we follow Cramer (1986) and implement a maximum of 30 iterations. Secondly, another frequent suggestion to address the local maximum problem is to repeatedly use different starting values for the numerical algorithm (Liu and Mahmassani (2000)) and to check whether the same solution is obtained. We adopt this idea and systematically change the vector of starting values within the boundaries of our parameter set \( \theta_k \) for the numerical algorithm by a Halton sequence based on the prime 7. Every time Stata reports successful convergence, we store the estimates and repeat this procedure using a new starting vector. The evaluation of \( \log L(\Delta_t(U_k|\theta_k)) \) is repeated 11 times and the estimates as well as the value of the likelihood function with the highest absolute value for \( \log L(\Delta_t(U_k|\theta_k)) \) are chosen.\(^{26}\)

5. Preliminary Analysis of the Results

In this and the subsequent section, we present the results from our simulation study. To begin with, we generated 100 hypothetical stocks by generating their time series of returns in Step 1 and estimated the market parameters in Step 2. Similar to Hey and Orme (1994), we estimate the utility functionals described in Table (1) based on the trading sequences generated in Step 3. The resulting number of investors, for which we evaluate the likelihood function \( L(\Delta_t(U_k|\theta_k)) \) is thus determined by the number of simulated stocks and the number of utility models considered. Accordingly, a total of 16 utility models have to be estimated as for each stock that we simulated in Step 1, one trading sequence per time series of returns for each utility-type investor has been generated, comprising 53 preference parameters each. As the evaluation process loops through all possible likelihood functions (one for each utility type) for each investor, the number of utility models to be evaluated sums up to a total of 16,000 utility models, requiring the estimation of 1,104,000 preference and nuisance parameters that need to be found numerically.\(^{25}\)

\(^{25}\)A trial-and-error search in terms of number of iterations and computational time shows that among the available numerical search techniques, the Berndt-Hall-Hall-Hausman algorithm (Berndt et al. (1974)) performs worst, which leaves the Newton-Ralphson and Davidon-Fletcher-Powell algorithm (Fletcher (1980)), a result that is in line with the results found by Griffiths et al. (1987). There is no clear winner between the latter two methods; thus, we compromise and use a mixed iteration procedure as described in the text. Note that if a quadratic approximation of \( L(\Delta_t(U_k|\theta_k)) \) provides a good description the log-likelihood function, then only a few steps suffice to find the maximum (if the fit is perfect, then the maximum can be found by one iteration only). We find that an acceptable quadratic approximation of \( L(\Delta_t(U_k|\theta_k)) \) prevails only for CRRA utility models (Train (2009)).

\(^{26}\)Note that there are plenty more ways to deal with local maxima: Exemplarily, a stochastic optimization algorithm that has the potential to overcome the local maximum trap was developed by Kirkpatrick et al. (1983), known as simulated annealing. Other techniques include gradient-sensitive hill climbing and random restart methods (see for details Rich and Knight (1991)).
Given the simulated time series of returns, this theoretically sums up to 40,320,000 single likelihood functions.²⁷

Before we begin our elaborations of our results, we need to identify potential problems that might interfere and compromise the quality of utility model selection. In particular, note that the quality and reliability of utility model selection procedures depend on the accuracy of the numerical evaluation of the respective utility models integrated in the likelihood function. Inspections of our results from the evaluation of \( \log L(\Delta_t(U_k|\theta_k)) \) revealed several problems associated with the evaluation of the likelihood function, as in particular, we detected missing values for \( \log L(\Delta_t(U_k|\theta_k)) \) associated with a stopped numerical search algorithm. Furthermore, we also found cases where values for \( \log L(\Delta_t(U_k|\theta_k)) \) and estimators for \( \theta_k \) are provided and numerical search reported convergence but standard errors were set to missing. Finally, we identified cases where we find large standard errors after the evaluation of \( \log L(\Delta_t(U_k|\theta_k)) \). We discuss each of these problems in detail as earlier simulation studies on utility model selection such as Carbone and Hey (1994) also explicitly report similar difficulties in the evaluation of the likelihood function but do not discuss their implications on their results in detail. Due to the fact that utility model selection based on \( AICC \) strongly depends on the surface of \( \log L(\Delta_t(U_k|\theta_k)) \), problems such as insufficient steepness of the likelihood function or convex segments can negatively affect the ranking of utility models and thus the accuracy of model selection (e.g. if the true utility model cannot be evaluated and is therefore not included in the model ranking as its \( AICC \) cannot be determined). While running the evaluation of the utility models \( k \) in Step 4, we noted that a number of utility models cannot be evaluated, notably because the iteration gets stuck or exceeds the maximum iteration steps such that the program provides missing values for \( \log L(\Delta_t(U_k|\hat{\theta}_k)) \). Before we started a more detailed investigation, we checked whether the true underlying utility model was assessed successfully in each evaluation and discard those cases where the true underlying utility model cannot be assessed as the corresponding likelihood function cannot be evaluated with respect to \( \theta_k \). This avoids biasing our findings when sorting the utility models according to their attained Akaike criterion. However, we did not discard those cases where Stata reports successful convergence and provides values for the likelihood function and \( \hat{\theta}_k|n,t \), but associated standard errors were set to missing. Table (2) provides an overview.

Since we run the calculations under a \( \lf \) specification of \( \log L(\Delta_t(U_k|\hat{\theta}_k)) \) in Stata, under which \( H(\Delta_t(U_k|\hat{\theta}_k)) \) and \( S(\Delta_t(U_k|\hat{\theta}_k)) \) are approximated numerically, we change the method and essential parts of the program to obtain the elements of \( H(\Delta_t(U_k|\hat{\theta}_k)) \) for a detailed analysis. In detail, we select those models where Stata reported missing values for standard errors or where the search algorithm failed to proceed after a number of iterations and rewrite the maximum likelihood program as a \( d2 \)-evaluator to gain further insight into the characteristics of the Hessian and the Information matrices. We find that for plausible values of \( \theta_k \), the determinant of the Hessian matrix \( \det H(\Delta_t(U_k|\hat{\theta}_k)) \) is indeed fairly

²⁷Technically, we estimate 176,000 models and 12,144,000 parameters as we repeat each estimation using different starting values as described in Step 4 (see Chapter 4).
pendent from the assumption that $H$ asymptotically equivalent to score vectors if the average scores were zero (Cramer (1986)). Note that this is product of the scores, which is an approximation of the covariance matrix of the tail, the Berndt-Hall-Hall-Hausman algorithm replaces Berndt-Hall-Hall-Hausman algorithm should overcome this by construction. In de-
mml model the algorithm (Berndt et al. (1974)) and additionally invoke the difficult
ods, such as for the Newton-Ralphson method, which runs into problems because
we refer to near-singularity if we mention singularity of $H$ is a real-valued variable and the precision of any statistical program is limited, thus in this paper
This has several consequences for some of the numerical methods, such as for the Newton-Ralphson method, which runs into problems because the stepsize, governed by $-H(\Delta_k(U_k|\theta_k))^{-1}$, cannot be determined if the Hessian matrix is degenerate. Indeed, we find terminations of the search algorithm predominantly in those iterations where the Newton-Ralphson method governs. It should be noted that the Davidon-Fletcher-Powell algorithm does not require an evaluation of $-H(\Delta_k(U_k|\theta_k))^{-1}$. To circumvent the impossibility of evaluating the inverse of $H(\Delta_k(U_k|\theta_k))$, we identify those models for which Newton-Ralphson algorithm failed to work, rerun the estimation with the Berndt-Hall-Hall-Hausman algorithm (Berndt et al. (1974)) and additionally invoke the difficult option at the nl model command. If the Hessian is singular for some values of $\theta_k$, then the Berndt-Hall-Hall-Hausman algorithm should overcome this by construction. In detail, the Berndt-Hall-Hall-Hausman algorithm replaces $H(\Delta_k(U_k|\theta_k))$ by the outer product of the scores, which is an approximation of the covariance matrix of the score vectors if the average scores were zero (Cramer (1986)). Note that this is asymptotically equivalent to $H(\Delta_k(U_k|\theta_k))$ (Griffiths et al. (1987)) and thus independent from the assumption that $H(\Delta_k(U_k|\theta_k))$ has full rank.\(^{29}\)

28 Technically, exact singularity of $H(\Delta_k(U_k|\theta_k))$ is difficult to obtain because the determinant is a real-valued variable and the precision of any statistical program is limited, thus in this paper we refer to near-singularity if we mention singularity of $H(\Delta_k(U_k|\theta_k))$. For completeness, a way to detect the extent of near-singularity has been outlined in Belsley et al. (2004).

29 Another possibility is to add a constant element to the diagonal elements of $H(\Delta_k(U_k|\theta_k))^{-1}$ until this expression becomes invertible. This is known as the Marquardt algorithm (Marquardt

Table 2. Frequency of Appearance for each Utility Model

<table>
<thead>
<tr>
<th></th>
<th>EUT</th>
<th>RDU</th>
<th>SPT</th>
<th>CPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>cal. %</td>
<td>% ~ calc. k % calc. % ~ calc. k % calc. % ~ calc. k % calc. % ~ calc. k</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>86.11%</td>
<td>1.52%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>QU82</td>
<td>0.00%</td>
<td>0.00%</td>
<td>73.99%</td>
<td>0.80%</td>
</tr>
<tr>
<td>KT92</td>
<td>0.00%</td>
<td>0.00%</td>
<td>67.77%</td>
<td>0.99%</td>
</tr>
<tr>
<td>None</td>
<td>83.90%</td>
<td>2.15%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>QU82</td>
<td>0.00%</td>
<td>0.00%</td>
<td>87.33%</td>
<td>0.26%</td>
</tr>
<tr>
<td>KT92</td>
<td>0.00%</td>
<td>0.00%</td>
<td>74.97%</td>
<td>1.30%</td>
</tr>
<tr>
<td>None</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>QU82</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>KT92</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

close to zero.\(^{28}\)
Table 3. Frequency of Appearance for each Utility Model for BHHH

The table captures the proportion of evaluated utility models to the total number of utility models evaluated if the Berndt-Hall-Hall-Hausman algorithm is used. The left column reports the share of utility model $k$ across all utility-type investors’ trading sequences, for which the corresponding likelihood function can be evaluated, denoted as $\%_{\text{calc.}}$. The right column captures the share of true utility model $k$, for which the corresponding likelihood function cannot be evaluated (i.e. where the numerical search algorithm was terminated) given the $k$-type investors’ trading sequence, denoted as $\%_{\neg\text{calc.}}$. Expected utility models are denoted as EUT, Rank-dependent Utility is denoted as RDU, Simple Prospect Theory (Kahneman and Tversky (1979)) uses the denotation SPT, whereas Cumulative Prospect Theory (Tversky and Kahneman (1992)) is denoted as CPT. Decision weights in accord to Quiggin (1982) are denoted as QU82 and as KT92 for decision weights as in Tversky and Kahneman (1992). If no decision weights are applicable, we used the abbreviation None. Furthermore, we use the denotation CRRA for CRRA utility functionals and EXPO to denote utility functions as in Saha (1993). For SPT and CPT, we use the denotation POWR for models with kinked power-functionals as in Kahneman and Tversky (1979) and used DHG0 to denote value functionals as defined in DeGiorgi and Hens (2006).

<table>
<thead>
<tr>
<th>Utility Model</th>
<th>EUT</th>
<th>RDU</th>
<th>SPT</th>
<th>CPT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% calc.</td>
<td>$%_{\neg\text{calc.}}$</td>
<td>calc.</td>
<td>$%_{\neg\text{calc.}}$</td>
</tr>
<tr>
<td>None</td>
<td>95.11%</td>
<td>0.62%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>QU82</td>
<td>0.00%</td>
<td>0.00%</td>
<td>90.99%</td>
<td>0.30%</td>
</tr>
<tr>
<td>KT92</td>
<td>0.00%</td>
<td>0.00%</td>
<td>89.17%</td>
<td>0.49%</td>
</tr>
<tr>
<td>None</td>
<td>99.0%</td>
<td>0.15%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>QU82</td>
<td>0.00%</td>
<td>0.00%</td>
<td>93.33%</td>
<td>0.16%</td>
</tr>
<tr>
<td>KT92</td>
<td>0.00%</td>
<td>0.00%</td>
<td>93.97%</td>
<td>0.31%</td>
</tr>
<tr>
<td>None</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>QU82</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>KT92</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>None</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>QU82</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>KT92</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

Re-estimation of these utility models using the Berndt-Hall-Hall-Hausman algorithm shows that we can cause the majority of them to converge, although a considerable sum of utility models still cannot be evaluated accurately (see Table (3)). This achievement comes with an additional computational burden, as more steps are required to provide a solution for $\hat{\theta}_k$, and estimators $\hat{\theta}_{k|n,t}$ differ significantly from the true parameters. Note that a large number of iterations is susceptible to yielding dubious results (Cramer (1986)), for likelihood functions with convex segments due to the incorporated decision-weighting function, as in RDU or due to the convex structure of some value functions as in SPT, this re-assessment clearly requires more than 30 iterations; however, it nevertheless yields ambiguous outcomes. The large number of iterations is consistent with the results of Griffiths et al. (1987), who find in their Monte Carlo simulation study that the Newton-Ralphson algorithm performs best under multicollinearity for a probit model, whereas the method of Berndt-Hall-Hall-Hausman is found to be least efficient. This is plausible, since the Berndt-Hall-Hall-Hausman algorithm is expected to work best if the likelihood (1963)); it represents the analogous case for a ridge regression (Cramer (1986)) that is usually recommended if multicollinearity is present.

30Application of the Berndt-Hall-Hall-Hausman algorithm provides several nuances with respect to its effectiveness; in particular, for CRRA, the algorithm yields results accompanied by an acceptable number of iteration steps, being not significantly different from the baseline procedure as outlined in the text, whereas the quality of $\hat{\theta}_{k|n,t}$ as well as the number of steps necessary deteriorates with RDU, SPT, and CPT, as well as with inclusion of heterogeneous error terms.
function can be sufficiently approximated by a second-order Taylor approximation (Train (2009)), but provides poor results in the face of highly non-linear likelihood functions (Pawitan (2001)). Consequently, literature from the field of experimental economics usually does not recommend the Berndt-Hall-Hall-Hausman method for estimation of utility functions, as pointed out by Harrison and Rutstrom (2008) and Harrison (2008).

Concerning the missing standard errors, the search algorithm declares convergence and a solution for $\hat{\theta}_k$ was found, although it cannot be considered to be reliable (Gould et al. (2006)). We find this problem especially prevalent for RDU, SPT, and CPT under decision weights according to Tversky and Kahneman (1992). Closer inspection of the iteration shows that Stata noted that the likelihood function is not concave in the last iteration step, indicating that the negative of the inverse of the Hessian matrix, which determines the stepsize of the numerical search algorithm, is not positive definite. Recall that if the likelihood function is concave, the Hessian matrix $H(\Delta_t(U_k|\hat{\theta}_k))$, the second derivative of the likelihood function, is negative definite for the full spectrum of $\hat{\theta}_k$, tantamount to a declining slope of log $L(\Delta_t(U_k|\hat{\theta}_k))$ with respect to one element of $\hat{\theta}_k$ ceteris paribus all other elements fixed. If $H(\Delta_t(U_k|\hat{\theta}_k))$ is negative definite, so is its inverse $H(\Delta_t(U_k|\hat{\theta}_k))^{-1}$ and the negative of the inverse $-H(\Delta_t(U_k|\hat{\theta}_k))^{-1}$, determining the stepsize in the Newton-Raphson method, is therefore positive definite. However, if $-H(\Delta_t(U_k|\hat{\theta}_k))^{-1}$ is negative definite, the Newton-Raphson algorithm moves down the slope of log $L(\Delta_t(U_k|\hat{\theta}_k))$ and thus away from the maximum. Rerunning the evaluation of the likelihood function using the Berndt-Hall-Hall-Hausman algorithm, we found that for the majority of our utility models standard errors are now provided. Note that the Berndt-Hall-Hall-Hausman algorithm moderates the problems of $-H(\Delta_t(U_k|\hat{\theta}_k))^{-1}$ as the direction of the search is determined by the outer product of the scores, which is necessarily positive definite. It therefore guarantees an increase of log $L(\Delta_t(U_k|\hat{\theta}_k))$ in each iteration, even if the log-likelihood function displays convex segments.

Furthermore, we also detected from some utility models, for which Stata reported successful convergence and where estimators for $\theta_k$ are provided, the associated standard errors, according to the Cramér-Rao Lower Bound (Rao (1945) and Cramer (1946)) derived from the inverse of $I(\Delta_t(U_k|\hat{\theta}_k))$, are large because the inverse of the Information matrix is small. Deficiencies in the likelihood function that result in inflated standard errors for $\hat{\theta}_{k|n,t}$ could indicate flat sections (such as plateaus and saddle points) in the surface of the likelihood function, probably stemming from $\Delta_t(U_k|\theta_k)$. Recall that large standard errors reflect the instability of $\hat{\theta}_{k|n,t}$ since a flat segment of log $L(\Delta_t(U_k|\hat{\theta}_k))$ contains theoretically infinitely many solutions of $\theta_k$. In particular, a flat surface of log $L(\Delta_t(U_k|\hat{\theta}_k))$ could point towards a certain degree of multicollinearity in log $L(\Delta_t(U_k|\hat{\theta}_k))$ (Griffiths et al. (1987)) or potential under-identification problems (Judge et al. (1985), Keele and Park (2006) and Greene (2008), for utility models see Carbone and Hey (1994)).
The effects of multicollinearity in linear models are well established (see exemplarily Judge et al. (1985), Lesaffre and Marx (1993), Belsley et al. (2004), Greene (2008) and Wooldridge (2010)),\textsuperscript{32} for nonlinear models such as the probit or logit however, the consequences are less certain, although some solutions have been suggested to transform the problem into a known linear one (e.g. Schaefer (1986)).\textsuperscript{33} This in turn suggests that the asymptotic properties of \( \hat{\theta}_{\text{NLT}} \) of the probit model under multicollinearity may also hold (McLeod (1974)) with consequences similar to the linear model. For a nonlinear likelihood model, multicollinearity can lead to dependencies between and within the score vectors \( S(\Delta_t(U_k|\theta_k)) \), thus invalidating the non-singularity assumption of the information matrix \( I(\Delta_t(U_k|\theta_k)) \) (Cramer (1986)). The likelihood function \( \log(L(\Delta_t(U_k|\theta_k))) \) then displays a ridge instead of a sharp peak, yielding inflated standard errors, instability in parameter estimates (as infinitely many solutions for \( \hat{\theta}_{\text{NLT}} \)), and shortcomings in the numerical search algorithms (Cramer (1986)), additionally compromising the precision of the estimates of \( \hat{\theta}_{\text{NLT}} \). Yet, it should be kept in mind that multicollinearity per se does not automatically invalidate the maximum likelihood properties, as simulation studies provide some evidence that the normal distribution property of the resulting distribution may still be intact (Griffiths et al. (1987)), especially if financial data contains a larger number of observations such that it can be expected that these particular asymptotic properties are likely to hold.

Ridges and plateaus in the likelihood function \( \log(L(\Delta_t(U_k|\theta_k))) \) both corresponding to singularity (or near-singularity) of the Hessian matrix \( H(\Delta_t(U_k|\theta_k)) \), could also point towards under-identification problems (Cramer (1986), Wooldridge (2010)), a problem also been detected in simulation studies similar to ours (Carbone and Hey (1994) and Carbone and Hey (1995)).\textsuperscript{34} The remedy for under-identification is to avoid side relations among the elements of \( \hat{\theta}_k \) similar to those reported by Gonzalez and Wu (1999). Note that if such an interdependence exists, then the variations of \( \theta_k \) and the estimated parameters \( \hat{\theta}_k \) are restricted to a subset of \( H(\Delta_t(U_k|\theta_k)) \) of less than the dimension of \( \theta_k \), denoted as \( K_k \). As a consequence, the rank of \( H(\Delta_t(U_k|\theta_k)) \) is less than \( K_k \) and some of the elements of \( \theta_k \) cannot be identified (Cramer (1986)). Consequently, the numerical search

\textsuperscript{31}Under moderate multicollinearity, the step size of a search algorithm is reduced if entering flat segments of \( \log(L(\Delta_t(U_k|\theta_k))) \) as a flattening of the likelihood function might indicate that the maximum is close (Train (2009)). If \( \log(L(\Delta_t(U_k|\theta_k))) \) is characterized by a flat surface over a large range of plausible \( \theta_k \) due to a sufficient degree of multicollinearity, the application of such an algorithm results in an increased number of iteration steps or a termination of the search procedure given a maximum number of iteration steps such that the respective utility model is not evaluated adequately.

\textsuperscript{32}For instance, in linear regression models, perfect multicollinearity leads to difficulties in inverting the vector product of the predictor variables (Belsley et al. (2004)).

\textsuperscript{33}In particular, Fomby et al. (1978) show for the linear probit model that by applying a principal component transformation, the information matrix \( I(\Delta_t(U_k|\theta_k)) \) can be restated in a known form, particularly as the inverse of the covariance matrix for weighted least squares (Griffiths et al. (1987)).

\textsuperscript{34}Multicollinearity and under-identification are difficult to disentangle and thus not widely discussed. We find this surprising since, despite their prevalence in studies on utility model selection and parameter estimation, under-identification issues are usually not discussed in detail as for the analysis, although a distinction is necessary to chose the remedies. If multicollinearity causes the breakdown in the likelihood estimation, collecting new data free from the defect can help. If a lack of identification is immanent to the models pretense, a modification of the model and a revision of extraneous information might be inevitable (Cramer (1986)).
algorithm endlessly cycles over the parameter space of $\theta_k$ as no unique maximand $\hat{\theta}_k$ of the likelihood function exists. Similar to multicollinearity, the fact that different admissible parameters of $\theta_k$ can define the same probability density, there exist infinitively many parameters that maximize $\log L(\Delta_t(U_k|\hat{\theta}_k))$. Side relations among the parameters of $\theta_k$ are partly immanent to utility models as, exemplarily for $RDU$, Yaari (1987) shows that even under risk neutrality, risk-averse behavior can be introduced by the nonlinearity in the decision weights. Rerunning our simulation and eliminating the error term $\epsilon$ in Step 3, we find that for a given trading sequence we can determine several parameter combinations within the spectrum of $\theta_k$ that can explain the observed trading sequence equally successful as we already conjectured in Chapter 3 where we discussed the role of the error term. Note that this observation is in line with experimental studies such as Gonzalez and Wu (1999), who report signs of significant correlation among their preference estimators for the $CPT$ case, and Carbone and Hey (1995), who mention that for $EUT$, they detect a large number of admissible parameter values that fit their data. For example, Keele and Park (2006) reported that the heteroscedastic linear probit model is quite prone to fragile identification (see also Judge et al. (1985) and Greene (2008)). He suggests this weakness to be evident if the likelihood function is smaller for the heteroscedastic likelihood than for the likelihood estimation where $\sigma_\epsilon$ is homogeneous and constrained to unity. We conclude that the magnitude of $\sigma_\epsilon$ as well as its stochastic properties are subject to discussion as perhaps a higher level of $\sigma_\epsilon$ or a different assumption of the distribution of the error term might capture the inherent problems of the likelihood function and the embedded utility functions therein more accurately. We discuss these points in the next chapter.\footnote{I am grateful to John Hey who pointed out the importance to run simulated trading sequences by adding an error term $\epsilon$ to dampen the problem of under-identification.}

6. Utility Model Selection: Analysis and Discussion

A first inspection of the results of model selection, depicted in Table (4), indicates that throughout the simulated trading sequences, the true utility models obtained the highest rank in more than 50% of all cases.\footnote{In detail, for $EUT$ being the true model, the correct utility model obtained on average in 55% of all cases the first rank, where for other utility models, the outcomes are better. For $RDU$, the true utility model obtained on average in 57.25% the first rank, for $SPT$ in 66% and for $CPT$ on average in 67.56% of all cases the first rank.} This is also reflected in the overall likelihood values across all utility models, which are close to zero for $EUT$ and $RDU$ and somewhat higher for $SPT$ and $CPT$.\footnote{The average of the likelihoods of $\log L(\Delta_t(U_{EUT}|\hat{\theta}_{EUT}))$ is $-6.42$, the average of $\log L(\Delta_t(U_{RDU}|\hat{\theta}_{RDU}))$ is $-8.22$, the average of $\log L(\Delta_t(U_{SPT}|\hat{\theta}_{SPT}))$ is $-22.87$ and for $\log L(\Delta_t(U_{CPT}|\hat{\theta}_{CPT}))$ is $-21.09$.} This is surprising as despite the obscure stochastic properties of the additional uncertainty stemming from $\Delta_t(U_k|\hat{\theta}_k)$, nuisance parameter $\hat{\sigma}_\epsilon$ seems to capture the stochastics from the additional uncertainty reasonably well. Although we noted that the estimated distribution of the error term $\hat{\sigma}_\epsilon = 0.031$ is on average higher than the distribution of the error term used in the simulation to generate the trading sequences, which we set to $\sigma_\epsilon = 0.01$, a t-test indicated that the estimators for $\sigma_\epsilon = 0.01$ are only significantly distinct from the true value on a 10%-significance level ($p$-value 0.088). Inspecting the estimators $\hat{\theta}_k$ of true utility models showed that the quality of our estimators varied among the different utility functions. We identified some cases, where the estimators are close to the true values (predominantly for $SPT$-type (83%) and $CPT$-type investors (78%)), in other cases the estimators are significant distinct even at a 1%-significance level (for more than 37% of all $EUT$-type
investors and for 43% of RDU-type investors). We will discuss the consequences of this unsharpeness of $\hat{\theta}_k$ in the subsequent analysis. Henceforth, these base case results serve as benchmark for our elaborations on the influence of modifications.

Table 4. Base Case Results: Ranking of each Utility Model

The table captures the median and average ranking (below), denoted as Rank of each utility model $k$ for trading sequences where utility model $k$ is the true utility model used to generate these trading sequences. We also report the type of the first-ranked (1st) and of the second-ranked (2nd) utility model (in brackets). Expected utility models are denoted as EUT, Rank-dependent Utility is denoted as RDU. Simple Prospect Theory (Kahneman and Tversky (1979)) uses the denotation SPT, whereas Cumulative Prospect Theory (Tversky and Kahneman (1992)) is denoted as CPT. Decision weights in accord to Quiggin (1982) are denoted as QU82, whereas Cumulative Prospect Theory (Tversky and Kahneman (1992)) for decision weights as in Tversky and Kahneman (1992). If no decision weights are applicable, we used the abbreviation None. Furthermore, we use the denotation CRRA (C) for CRRA utility functionals and EXPO (E) to denote utility functions as in Saha (1993). For SPT and CPT, we use the denotation POWR (P) for models with kinked power-functionals as in Kahneman and Tversky (1979) and used DGH0 (D) to denote value functionals as defined in DeGiorgi and Hens (2006). Significance levels are calculated according to Wilcoxon signed-rank tests comparing the average ranking between the first-ranked and second-ranked utility model. We use * *, **, and *** for significance at the 1%, 5% and 10% levels respectively.

<table>
<thead>
<tr>
<th>EUT</th>
<th>RDU</th>
<th>SPT</th>
<th>CPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank</td>
<td>Rank</td>
<td>Rank</td>
<td>Rank</td>
</tr>
<tr>
<td>1st</td>
<td>2nd</td>
<td>1st</td>
<td>2nd</td>
</tr>
<tr>
<td>None</td>
<td>1***</td>
<td>1.569</td>
<td>EUT$^C$</td>
</tr>
<tr>
<td>QU82</td>
<td>0</td>
<td>1***</td>
<td>RDU$^E$</td>
</tr>
<tr>
<td>KT92</td>
<td>0</td>
<td>1***</td>
<td>RDU$^E$</td>
</tr>
<tr>
<td>None</td>
<td>1***</td>
<td>1.510</td>
<td>EUT$^E$</td>
</tr>
<tr>
<td>QU82</td>
<td>0</td>
<td>1***</td>
<td>RDU$^E$</td>
</tr>
<tr>
<td>KT92</td>
<td>0</td>
<td>1***</td>
<td>RDU$^E$</td>
</tr>
<tr>
<td>QU82</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>KT92</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>None</td>
<td>1***</td>
<td>1.490</td>
<td>RDU$^E$</td>
</tr>
<tr>
<td>QU82</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>KT92</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

We pointed out earlier that our simulation was conducted using a normally distributed error term $\epsilon$, arbitrarily setting its standard deviation to $\sigma_\epsilon = 0.01$. However, in the light of the significant deviations in $\hat{\theta}_k$ particularly for EUR and RDU, the magnitude of $\sigma_\epsilon$ might be too low. Recall that with regard to the significance of our parameter estimates of each utility model $k$, we found that the introduction of such an error term can help to resolve parts of the under-identification problems. To elaborate the effects of $\epsilon$, consider a situation in which the market parameters
In empirical data, however, it cannot be expected that the error term plays such an important role with respect to the magnitude of the error term, we rerun our simulation for various values of the standard deviation $\sigma$.

To investigate the sensitivity of our utility model selection, we alter the trading sequences and thus affect utility model selection negatively. To locate the source of these additional uncertainties, we rerun simulations with a stationary stochastic process using an Ornstein-Uhlenbeck Process (see Ornstein and Uhlenbeck (1930), Box et al. (2015)), generating almost time-invariant estimators for $R_{U,t}$, $R_{D,t}$, and for which we fixed $p_t = 50\%$. If additional disturbance arises from the estimation of market parameters, then we expect a notable improvement in $\log L(\Delta_t(U_k|\theta_k))$.

Note that as a modification of the stochastic process affects all utility models, if additional uncertainty caused by market parameter estimation, all likelihoods are expected to improve significantly. Inspections of $\log L(\Delta_t(U_k|\theta_k))$ for the different utility models showed that for EUT the maximized likelihood is close to zero, whereas for SPT and CPT models, the overall likelihood remains large, although still ranging in lower double figures, all not significantly distinct from the baseline likelihood values on a 10%-significance level according to a likelihood-ratio test. The ranking of all utility models compared to the baseline case in Table (4) shows no significant changes according to a Wilcoxon signed-rank test. We detected only little improvement in $\log L(\Delta_t(U_k|\theta_k))$ for all utility models according to likelihood ratio tests for usual significance levels. We conclude that the source of additional disturbance stems from accrued gains and losses in terms of $W_t$.

Tracing back the additional uncertainty to the dynamics of $W_t$, it is not clear, whether the estimation of the nuisance parameter $\sigma_\epsilon$ can capture this additional uncertainty stemming from $\Delta_t(U_k|\theta_k)$, particularly if we try to overlay the defects from $\Delta_t(U_k|\theta_k)$ by increasing $\sigma_\epsilon$. We observed that an increase in $\sigma_\epsilon$ alters the trading sequences and thus the magnitude of $W_t$ within the likelihood function, however, introducing higher disturbance of the error term could lead to a deterioration of the estimators, alters the trading sequences and thus affect utility model selection negatively. To investigate the sensitivity of our utility model selection with respect to the magnitude of the error term, we rerun our simulation for various values of the standard deviation $\sigma_\epsilon$. According to Table (5), a more pronounced standard deviation of the error term $\sigma_\epsilon$ is detrimental to utility model selection. In empirical data however, it cannot be expected that the error term plays such a role.

--

38The significance of the differences in the rank scores between the baseline setting and the results from a re-estimation under modified settings are based on a Wilcoxon signed-rank test (Wilcoxon (1945), for the two-sample test see Mann and Whitney (1947)). This non-parametric tests seems appropriate for this purpose as it can be used for outcomes that are coded ordinally and that requires no explicit distribution of the matched samples, yet the test is comparable to a t-test (Siegel (1956)).
Table 5. Variations of Error Term: Ranking of each Utility Model

The table captures the median and average ranking (below) of each utility model \( k \) for trading sequences where utility model \( k \) is the true utility model used to generate these trading sequences given various values of \( \sigma \). Expected utility models are denoted as \textit{EUT}, Rank-dependent Utility as denoted as \textit{RDU}. Simple Prospect Theory (Kahneman and Tversky (1979)) uses the denotation \textit{SPT}, whereas Cumulative Prospect Theory (Tversky and Kahneman (1992)) is denoted as \textit{CPT}. Decision weights in accord to Quiggin (1982) are denoted as \textit{QU82} and as \textit{KT92} \( (K) \) for decision weights as in Tversky and Kahneman (1992). If no decision weights are applicable, we used the abbreviation \textit{None}. Furthermore, we use the denotation \textit{CRRA} for CRRA utility functionals and \textit{EXPO} to denote utility functions as in Saha (1993). For \textit{SPT} and \textit{CPT}, we use the denotation \textit{POWR} for models with kinked power-functionals as in Kahneman and Tversky (1979) and used \textit{DHG0} to denote value functionals as defined in DeGiorgi and Hens (2006). Significance levels are calculated according to Wilcoxon signed-rank tests comparing the average ranking between the first-ranked utility model if the true utility model obtains first rank and to the first-ranked model if the true model obtains second rank. For the respective utility models of first- and second rank we refer to Table (7). We use \textit{***}, \textit{**} and \textit{*} for significance at the 1%, 5% and 10% levels respectively.

<table>
<thead>
<tr>
<th></th>
<th>EUT</th>
<th>RDU</th>
<th>SPT</th>
<th>CPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma = 0.05 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{None}</td>
<td>1***</td>
<td>2**</td>
<td>2**</td>
<td>0</td>
</tr>
<tr>
<td>\textit{QU82}</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1**</td>
</tr>
<tr>
<td>\textit{KT92}</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>\textit{EUT}</td>
<td>1.630</td>
<td>1.710</td>
<td>1.790</td>
<td>-</td>
</tr>
<tr>
<td>\textit{RDU}</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>\textit{SPT}</td>
<td>1.570</td>
<td>1.680</td>
<td>1.750</td>
<td>1.580</td>
</tr>
<tr>
<td>\textit{CPT}</td>
<td>1.700</td>
<td>1.770</td>
<td>1.620</td>
<td>1.720</td>
</tr>
<tr>
<td>( \sigma = 0.10 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \sigma = 0.20 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \sigma = 0.05 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{None}</td>
<td>1***</td>
<td>2**</td>
<td>2**</td>
<td>0</td>
</tr>
<tr>
<td>\textit{QU82}</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>\textit{KT92}</td>
<td>0</td>
<td>0</td>
<td>1**</td>
<td>2</td>
</tr>
<tr>
<td>\textit{EUT}</td>
<td>1.610</td>
<td>1.770</td>
<td>1.750</td>
<td>-</td>
</tr>
<tr>
<td>\textit{RDU}</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>\textit{SPT}</td>
<td>1.600</td>
<td>1.700</td>
<td>1.770</td>
<td>-</td>
</tr>
<tr>
<td>\textit{CPT}</td>
<td>1.690</td>
<td>1.800</td>
<td>1.870</td>
<td>1.530</td>
</tr>
<tr>
<td>( \sigma = 0.10 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \sigma = 0.20 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \sigma = 0.05 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{None}</td>
<td>1***</td>
<td>2**</td>
<td>2**</td>
<td>0</td>
</tr>
<tr>
<td>\textit{QU82}</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>\textit{KT92}</td>
<td>0</td>
<td>0</td>
<td>1**</td>
<td>2</td>
</tr>
<tr>
<td>\textit{EUT}</td>
<td>1.650</td>
<td>1.660</td>
<td>1.520</td>
<td>-</td>
</tr>
<tr>
<td>\textit{RDU}</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>\textit{SPT}</td>
<td>1.490</td>
<td>1.440</td>
<td>1.550</td>
<td>1.700</td>
</tr>
<tr>
<td>\textit{CPT}</td>
<td>1.610</td>
<td>1.790</td>
<td>1.620</td>
<td>1.720</td>
</tr>
<tr>
<td>( \sigma = 0.10 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \sigma = 0.20 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \sigma = 0.05 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{None}</td>
<td>1***</td>
<td>2**</td>
<td>2**</td>
<td>0</td>
</tr>
<tr>
<td>\textit{QU82}</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>\textit{KT92}</td>
<td>0</td>
<td>0</td>
<td>1**</td>
<td>2</td>
</tr>
<tr>
<td>\textit{EUT}</td>
<td>1.530</td>
<td>1.650</td>
<td>1.730</td>
<td>-</td>
</tr>
<tr>
<td>\textit{RDU}</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>\textit{SPT}</td>
<td>1.630</td>
<td>1.740</td>
<td>1.630</td>
<td>1.740</td>
</tr>
<tr>
<td>\textit{CPT}</td>
<td>1.700</td>
<td>1.840</td>
<td>1.700</td>
<td>1.840</td>
</tr>
</tbody>
</table>
modest role as in our simulation. Thus, we wish to eliminate the possibility that the likelihood approach shows a tendency to prefer a particular utility model if the error term dominates (i.e. other trading factors matter and preference considerations are negligible).

**Figure 2. Distribution of** $\log L(\Delta_t(U_k|\hat{\theta}_k))$ **for different Investors**

The figures on the left exemplarily illustrate the distribution of the log-likelihood values $\log L(\Delta_t(U_{EUT,CRRA}|\hat{\theta}_{EUT,CRRA}))$, divided by the number of simulated draws for the trading sequence of a EUT-type investor with CRRA utility functional. The figure on the right shows the distribution of the log-likelihood values $\log L(\Delta_t(U_{EUT,CRRA}|\hat{\theta}_{EUT,CRRA}))$ divided by the number of simulated draws of the EUT model with CRRA utility functional for a trading sequence of a Random Trader where $\epsilon \sim N(0, 1)$.

To investigate the sensitivity of our utility model selection with regard to a further increase in $\sigma_\epsilon$, we extend our simulation and introduce the concept of a Random Trader, defined as an investor who trades based on economically irrelevant criteria that are purely independent from preferences (Kyle (1985) and Black (1986)). For our simulation of the trading sequences, we define a Random Trader as an investor, who trades on other criteria that are unsystematic, standardized in their variance, independent from utility considerations and thus approximately normally distributed with $\epsilon \sim N(0, 1)$. We make the conjecture that the respective log-likelihood values $\log L(\Delta_t(U_k|\hat{\theta}_k))$ of each of the $k$ utility models should be close to the baseline log-likelihood $\log L(\Delta_t(\epsilon)) = -174.67$ for a random traders’ trading sequence, as none of the models contributes further information to observed trading data. In particular, to construct a random trader, we generate trading signals by replacing the difference in utilities $\Delta_t(U_k|\hat{\theta}_k)$ by a stochastic element $\epsilon$, characterized by a standard normal density with zero mean and standardized variance as mentioned above. Accordingly, the random trader has a positive exposure in the risky stock if argument $\epsilon$ yields a cumulative density $\Phi(\epsilon)$ above 50% and the investor prefers to hold the riskless investment otherwise. In all cases, the obtained values of $\log L(\Delta_t(U_k|\hat{\theta}_k))$ are not significantly different from the baseline log-likelihood $\log L(\Delta_t(\epsilon))$ according to likelihood-ratio tests conducted for each utility model $k$.\(^{39}\) The differences in the likelihood log $\log L(\Delta_t(U_k|\hat{\theta}_k))$ for the EUT-type investors’ trading sequence and the likelihood of a Random Traders’ trading sequence is also illustrated in Figure 2. The log-likelihood of the simulated EUT-type investor is close to its theoretical maximum, whereas in contrast, the log-likelihood values of the same utility model given a trading sequence of a Random Trader are distributed

\(^{39}\)The corresponding $p$-values range from $p$-value 0.289 for SPT given POWR value functional and a decision weight according to Quiggin (1982) up to $p$-value 0.382 for EUT given EXOP utility functional.
around ln(0.5) and displays a higher dispersion.

Given the trading sequence of a Random Trader, sorting the contemplable utility models according to the AICC reveals a pronounced tendency to identify an SPT model specification as the best fit, regardless of the combination of the various return moment characterizations and stochastic processes (Wilcoxon signed-rank test p-value 0.054). We also find it noteworthy that this effect persists if we simulate the trading behavior with a certain degree of dispersion by increasing the magnitude of \( \sigma \) by using \( \epsilon \sim N(0, \sigma^2) \). In particular, we consider two cases in which we trigger a hold signal whenever the cumulative density \( \Phi(\Delta_t(U_k|\theta_k)/\sigma(\Delta_t(U_k|\theta_k))) > 50\% \), where the spread of the cumulative normal distribution depends on the difference in utility, as is frequently assumed in experiments (Moffatt and Peters (2001) and Loomes et al. (2002)). Note that this form of the error term corresponds to cases where an investor makes more mistakes the larger the difference in utility. While we obtain acceptable quality of model selection results for the endogenous case, we detect a bias toward SPT in the latter case, notably for particularly high values of \( \sigma(\epsilon) \), as the resulting trading behavior resembles that of the random investor.\(^{40}\) This may indicate some problems in identifying the correct utility model specification if differences in utility represent only a minor aspect for investor decision making in stock markets, as Grinblatt and Keloharju (2001c) and Levitt and List (2007) point out.

As a central component of the decision model (3.2), the role of the stochastic error term is crucial to determination of the likelihood function as we showed, however, the correct formal specification of \( \epsilon \) is the subject of an ongoing debate (see Harless and Camerer (1994), Hey and Orme (1994), Loomes and Sugden (1995) and Ballinger and Wilcox (1997) for a discussion of Cauchy and Laplace distributed errors), since the distributional assumption may have consequences on the determination of the best-fitting model. A frequently applied form is to define a utility ratio index, discussed in Harrison and Rutstrom (2008), related with a logit formulation as shown by McFadden (1974). To test the sensitivity of our results regarding a modification of the distribution of the error term, we rerun our simulations under an independently, identically extreme value (type 1) distributed error term \( \epsilon \) with density \( \phi(\epsilon) \sim \frac{1}{\pi \sigma} e^{-\epsilon/\sigma} \). The cumulative distribution of the error term used for the likelihood function is therefore \( \Phi(\epsilon) = e^{-\epsilon/\sigma} \) and the difference in utility plus the error term is described by a logistic function (Greene (2008), Train (2009) and Hosmer et al. (2013), for utility model selection see Carbone and Hey (1995), Loomes and Sugden (1995), Loomes et al. (2002) and Harrison (2008)).\(^{41}\) According

\(^{40}\)We noted that the length of the trading sequences from the Random Trader is sensitive to the mean of \( \epsilon \) and insensitive with respect to \( \sigma \). The resulting trading sequences are short if the mean of \( \epsilon \) is set to zero, comparable to a trading pattern from a daytrader. If longer trading sequences are required, the mean of \( \epsilon \) should be close to 0.013 to match with round-trip length of an EUT-type investor in our case. Due to the insensitivity regarding \( \sigma \), the corresponding mean for \( \epsilon \) can be approximated by matching the upper bound of the integral of the cumulative distribution of the Random Trader with the investment ratio of an EUT-type investor (number of observations where the indicator is one divided by trading days). We estimate a proportional hazard model as proposed by Cox (1972) to verify the conjecture that the generated trading sequences of the random investor are akin to the sequences generated by introducing the disturbance \( \sigma(\Delta_t(U_k|\theta_k)) \).

A log-rank test (Mantel (1966)), as implemented in the \texttt{sts test} command in Stata, reveals no significant difference between the two hazard rates.

\(^{41}\)Note that the mean of the extreme value distribution is not zero as for the standard normal distribution (the usual interpretation of a zero mean is that investors don’t do errors on average, see Carbone and Hey (1995)), however, the mean appears to be immaterial.
to Table (6) we found our results to be virtually unchanged. A Wilcoxon signed-rank test indicates that the changes in the average rank between the baseline group and the average ranks, where we used a logit specification of the error term instead, are insignificant.\footnote{Exemplarily, for \textit{EUT}, the corresponding \textit{p}-value is 0.504, for \textit{RDU} its \textit{p}-value is 0.373, for \textit{SPT} and \textit{CPT} the corresponding \textit{p}-values are 0.324 and 0.425, respectively.} Although the logistic distribution is characterized by fatter tails in comparison to the standard normal distribution, the differences between these distributions are usually insignificant (Hosmer et al. (2013)), which explains that we find virtually the same results.\footnote{However, despite this result, we suggest not to use logit for several reasons: First, we cannot exclude correlation effects in the error term as Train (2009) points out. If these correlations exist, they need to be modeled and estimated explicitly (Train (1986) and Train (2009)). Second, for utility models where the value of the utility can be negative, such as \textit{SPT} and \textit{CPT}, this concept violates certain axioms of rationality, since a logit specification is justified only under a positive measure scale such as e.g. \textit{EUT} (see Luce (1959)).} Despite logit and probit, various other error specifications and their relation to utility specifications are suggested and reviewed in Hey (1995), Loomes and Sugden (1995), Ballinger and Wilcox (1997), Hey (2002) and Loomes et al. (2002). Note that, under certain circumstances, using extreme value distributed error terms can bias the results: Wilcox (2008) shows that in a logit model according to Luce (1959), the finding that subjects behave according to \textit{IARA} may be biased by the fact that \(\epsilon\) follows an extreme value distribution.\footnote{In the experimental literature, the heterogeneous error likelihood specification is also referred to as \textit{Fechnerian Error} or \textit{white noise} (Fechner (1966), a choice model has been developed by Becker and Marschak (1963) and popularized by Hey and Orme (1994)), consistent with a probit specification, as shown above.} Although the logistic distribution is characterized by fatter tails in comparison to the standard normal distribution, the differences between these distributions are usually insignificant (Hosmer et al. (2013)), which explains that we find virtually the same results.

With regard to the specification and estimation of \(\sigma_\epsilon\), we mention that the nuisance parameter may also be able to cope with other issues arising with financial data such as the correlation structure within \(\sigma_\epsilon\), i.e. error in decision making of an investor carries over from previous periods, creating correlation among \(\epsilon\). Autocorrelation within the structure of the error term may arise if other factors affecting trading decisions, that are assumed to be independent from individual risk preferences, correlate across time, which is suspected to alter the dispersion of \(L(\Delta_t(U_k|\theta_k))\) and thus should be captured by \(\sigma_\epsilon\). Since autocorrelation does not affect the functional form of \(L(\Delta_t(U_k|\theta_k))\) per se, despite increasing its spread via its dispersion (Pawitan (2001)), estimating \(\sigma_\epsilon\) should have no consequences for the selection of the utility model, but it can indirectly interact with model identification via the precision of estimates \(\hat{\theta}_{k|n,t}\).\footnote{Note that the estimation of parameter \(\sigma_\epsilon\) introduces an element that lowers the concavity of \(L(\Delta_t(U_k|\theta_k))\) and increases the number of iteration steps of our numerical search algorithm.} To investigate the impact of autocorrelation on \(\sigma_\epsilon\), we consider the possibility that the error term \(\epsilon\) may be potentially autocorrelated within each investor’s trading sequence in reality and that such correlation may drive the high standard errors we expect to find in this case (Wooldridge (2010)). Accordingly, in an earlier pretest of our likelihood approach, we modeled a lagged error term with lag one to analyze its effect on our utility model selection results. We follow Roger (1994) and use the \texttt{cluster} option in Stata within the \texttt{ml model} command, which invokes evaluation of the likelihood function (see Harrison and Rutstrom (2008) and Harrison (2008)) to control for the autocorrelation structure within each investor type (Harrison and Rutstrom (2009)). We find that this action leads to virtually no consequences for both our results and the precision of the estimated risk parameters, such that we conclude that \(\hat{\sigma}_\epsilon\) is also able to capture possible autocorrelation sufficiently.

\textit{42}\footnote{Exemplarily, for \textit{EUT}, the corresponding \textit{p}-value is 0.504, for \textit{RDU} its \textit{p}-value is 0.373, for \textit{SPT} and \textit{CPT} the corresponding \textit{p}-values are 0.324 and 0.425, respectively.} Exemplarily, for \textit{EUT}, the corresponding \textit{p}-value is 0.504, for \textit{RDU} its \textit{p}-value is 0.373, for \textit{SPT} and \textit{CPT} the corresponding \textit{p}-values are 0.324 and 0.425, respectively.\textit{43} However, despite this result, we suggest not to use logit for several reasons: First, we cannot exclude correlation effects in the error term as Train (2009) points out. If these correlations exist, they need to be modeled and estimated explicitly (Train (1986) and Train (2009)). Second, for utility models where the value of the utility can be negative, such as \textit{SPT} and \textit{CPT}, this concept violates certain axioms of rationality, since a logit specification is justified only under a positive measure scale such as e.g. \textit{EUT} (see Luce (1959)).\textit{44} In the experimental literature, the heterogeneous error likelihood specification is also referred to as \textit{Fechnerian Error} or \textit{white noise} (Fechner (1966), a choice model has been developed by Becker and Marschak (1963) and popularized by Hey and Orme (1994)), consistent with a probit specification, as shown above.\textit{45} Note that the estimation of parameter \(\sigma_\epsilon\) introduces an element that lowers the concavity of \(L(\Delta_t(U_k|\theta_k))\) and increases the number of iteration steps of our numerical search algorithm.
Table 6. Logit Error Specification: Ranking of each Utility Model

The table captures the median and average ranking (below), denoted as Rank of each utility model $k$ for trading sequences where utility model $k$ is the true utility model used to generate these trading sequences given a logit specification of the error term. We also report the type of the first-ranked (1st) and of the second-ranked (2nd) utility model (in brackets). Expected utility models are denoted as EUT, Rank-dependent Utility is denoted as RDU, Simple Prospect Theory (Kahneman and Tversky (1979)) uses the denotation SPT, whereas Cumulative Prospect Theory (Tversky and Kahneman (1992)) is denoted as CPT. Decision weights in accord to Quiggin (1982) are denoted as QU82 ($Q$) and as KT92 ($K$) for decision weights as in Tversky and Kahneman (1992). If no decision weights are applicable, we used the abbreviation None. Furthermore, we use the denotation CRRA (C) for CRRA utility functionals and EXPO (E) to denote utility functions as in Saha (1993). For SPT and CPT, we use the denotation POWR (P) for models with kinked power-functionals as in Kahneman and Tversky (1979) and used DHG0 (D) to denote value functionals as defined in DeGiorgi and Hens (2006). Significance levels are calculated according to Wilcoxon signed-rank tests comparing the average ranking between the first-ranked and second-ranked utility model. We use ***, ** and * for significance at the 1%, 5% and 10% levels respectively.

<table>
<thead>
<tr>
<th>CRRA</th>
<th>EUT</th>
<th>RDU</th>
<th>SPT</th>
<th>CPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>1***</td>
<td>EUTC</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.560</td>
<td>EUTC</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>QU82</td>
<td>0</td>
<td>1***</td>
<td>RDU</td>
<td>1***</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>1.550</td>
<td>(RDU)</td>
<td>1.420</td>
</tr>
<tr>
<td>KT92</td>
<td>0</td>
<td>1***</td>
<td>RDU</td>
<td>1***</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>1.500</td>
<td>(RDU)</td>
<td>1.480</td>
</tr>
<tr>
<td>EXPO</td>
<td>None</td>
<td>1***</td>
<td>EUTC</td>
<td>0</td>
</tr>
<tr>
<td>1.510</td>
<td>EUTC</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>QU82</td>
<td>0</td>
<td>1***</td>
<td>RDU</td>
<td>0</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>1.510</td>
<td>(RDU)</td>
<td>-</td>
</tr>
<tr>
<td>KT92</td>
<td>0</td>
<td>1***</td>
<td>RDU</td>
<td>0</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>1.490</td>
<td>(RDU)</td>
<td>-</td>
</tr>
<tr>
<td>POWER</td>
<td>QU82</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.480</td>
</tr>
<tr>
<td>KT92</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.510</td>
</tr>
<tr>
<td>DGH0</td>
<td>QU82</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.530</td>
</tr>
<tr>
<td>KT92</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.420</td>
</tr>
</tbody>
</table>

Other issues regarding the quality of utility model selection concern the likelihood function $\log L(\Delta_k(U_k(\theta_k)))$ and its susceptibility of overfitting issues. Although we control for the different number of parameters by using AICC instead of $\log L(\Delta_k(U_k(\theta_k)))$, we cannot be absolutely sure whether the penalization term for additional parameters is sufficient. The inclusion of a piecewise negative exponential value functional as proposed by DeGiorgi and Hens (2006), which contains four different risk sensitivity parameters, allows to test for the susceptibility of the likelihood based model selection procedure for possible overfitting issues.\footnote{These authors mention that this functional form should offer a higher descriptive power with respect to the behavior of investors in financial markets than other functions, since this form is explicitly designed to capture the features of decreasing marginal utility if financial outcomes}
proposed value function added to our analysis is accordingly defined as a piece-wise negative exponential value function with parameters set equal to the values presented in DeGiorgi and Hens (2006) to match with parameter estimates of Tversky and Kahneman (1992). If the likelihood function is prone to overfitting, then models containing a DeGiorgi and Hens (2006) functional should end up in higher ranks compared to models incorporating Kahneman and Tversky (1979) versions of the value function. We find no signs of an amassment of DGH0 models in higher ranks across all utility models, thus an overfitting problem does not appear to be significant.\footnote{In those cases where the DGH0-type is not the true model, the average rank for DGH0 utility model ranges between 12.4 (median rank: 11) for EUT and 7.2 (median rank: 7) for SPT. P-values from the Wilcoxon signed-rank tests range from 0.024 for EUT up to 0.051 for SPT, thus the average ranks of the two models are at least significantly distinct on a 10%-significance level.}

Since some utility models, for which we generate trading sequences in Step 3, are nested in more general formulations, we expect that nesting utility models should yield estimates for $\hat{\theta}_k$ statistically indistinguishably close to those constraining values under which the nested utility model coincides with a nesting one. We find that for an increase in $\sigma$, the trading sequences shorten and our parameter estimates for $\theta_k$ deteriorate, thus due to the high standard errors and the imprecision in the preference parameter estimates, nesting models are preferred to nested models even in those cases where nested models represent the true underlying utility model (see Table (7)). We elaborate above that our simulations reveal signs of shortcomings in the functionality of the applied search algorithm and the reliability of $\hat{\theta}_k|_{n,t}$. These drawbacks compromise not only the obtained likelihood maxima and the information criteria, based upon which the ranking of models takes place, but they also affect the precision of our estimators. If we expect some deviations from the implicitly inherent constraining parameter constellations in terms of $\hat{\theta}_k$, then we obtain a ranking of utility models where nesting models should prevail in the upper ranks.

A closer inspection of Table (7) for example reveals for an increase in the dispersion of the error term (increase in $\sigma_e$) a tendency towards RDU where EUT is the correct underlying model. Likewise, we detect a tendency to select exponential power functionals over power functionals, which implies that at the scaling parameter $\rho$ in the EXPO utility functional is significantly distinct from zero - the case where EXPO coincides with CRRA.\footnote{It should be kept in mind that the correct specification is a CRRA utility function such that $\log L(\Delta_{t}(U_{EUT}(\hat{\theta}_{EUT}))$ is theoretically close to zero. We model the expected utility case explicitly using CRRA as a ranking tendency toward expo-power utility, which may strengthen our conjecture that nesting models are systematically preferred. Inspecting the log-likelihood values and the information criterion of our simulations, the best fitting model reveals an AICC indeed fairly close to zero (a small differential is due to sample size and parameter correction), a result which is independent from the period for which we calculate returns, which is a result of the inherent horizon independence of CRRA utility models (Merton (1969)).} We find that in those cases where the expo-power function obtains the top rank, the scaling parameter $\rho$ is significantly different from zero according to a t-test ($p$-value 0.031), an indication that the expo-power function does not coincide with a CRRA utility, although we would reach the edges of the return distributions. They argue that marginal utility stemming from the value function is still decreasing at the bounds of the return distribution, whereas the usually applied form of Kahneman and Tversky (1979) is virtually linear in the realm of higher stakes.
Table 7. Variations of Error Term: Tendency towards Nesting Utility Models

The table reports the type of the first-ranked (1\textsuperscript{st}) and of the second-ranked (2\textsuperscript{nd}) utility model (in brackets) for various values of $\sigma$. Expected utility models are denoted as EUT, Rank-dependent Utility is denoted as RDU. Simple Prospect Theory (Kahneman and Tversky (1979)) uses the denotation SPT, whereas Cumulative Prospect Theory (Tversky and Kahneman (1992)) is denoted as CPT. Decision weights in accord to Quiggin (1982) are denoted as QU\textsubscript{82} (Q) and as KT92 (K) for decision weights as in Tversky and Kahneman (1992). If no decision weights are applicable, we used the abbreviation None. Furthermore, we use the denotation CRRA (C) for CRRA utility functionals and EXP (E) to denote utility functions as in Saha (1993). For SPT and CPT, we use the denotation POWER (P) for models with kinked power-functionals as in Kahneman and Tversky (1979) and used DGH0 (D) to denote value functionals as defined in DeGiorgi and Hens (2006).

<table>
<thead>
<tr>
<th></th>
<th>EUT</th>
<th>RDU</th>
<th>SPT</th>
<th>CPT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\sigma = 0.05$</td>
<td>$\sigma = 0.10$</td>
<td>$\sigma = 0.20$</td>
<td>$\sigma = 0.05$</td>
</tr>
<tr>
<td>None</td>
<td>EUT\textsuperscript{C} (EUT\textsuperscript{E})</td>
<td>RDU\textsuperscript{E} (RDU\textsuperscript{E})</td>
<td>SPT\textsuperscript{E}</td>
<td>CPT\textsuperscript{E}</td>
</tr>
<tr>
<td>QU\textsubscript{82}</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>KT92</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>None</td>
<td>EUT\textsuperscript{E} (EUT\textsuperscript{E})</td>
<td>SPT\textsuperscript{E} (RDU\textsuperscript{E})</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>QU\textsubscript{82}</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>KT92</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>None</td>
<td>POWER (POWER)</td>
<td>SPT\textsuperscript{E} (RDU\textsuperscript{E})</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>QU\textsubscript{82}</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>KT92</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>None</td>
<td>DGH0 (DGH0)</td>
<td>SPT\textsuperscript{E} (RDU\textsuperscript{E})</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>QU\textsubscript{82}</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>KT92</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
expect $\rho$ not to be significantly distinct from zero. The tendency to rank nesting utility models to nested models if the true underlying round-trip sequence is determined by a nested utility model also carries over to the case of non-expected utility investor trading decisions. For instance, if the decision process is determined by RDU preferences with a CRRA value function and decision weights according to (A.6), we detect significant differences of $\theta_{\text{RDU}}$ from the true parameterization $\theta_{\text{RDU}}$. Exemplarily, for EUT, we expect $\gamma$ to be close to one as EUT and RDU coincide if $\gamma = 1$ and for RDU, parameter $\gamma$ should range near 0.65, however, in cases where EUT is the true utility model but RDU obtains the highest rank, $\gamma$ is near one (mean value for $\gamma$ is 0.92), but the hypothesis that $\gamma$ is one, is rejected at a 5%-significance level ($p$-value 0.049 according to an independent t-test). As the decision weight $\omega(\hat{p}_{j,t}|\hat{\theta}_{\text{RDU}})$ seems to capture parts of the risk aversion of the EUT model as estimates for $\delta$ are close to risk neutrality ($p$-value 0.122), given the additional uncertainty stemming from the difference in utility, which is in line with Yaari (1965). Consequently, $\log(L(\Delta_t(U_{\text{RDU}}|\hat{\theta}_{\text{RDU}})))$ obtains higher values than the corresponding likelihood function of EUT although EUT is the true underlying utility model. In conclusion, the likelihood measure seems to identify models that are close to the true model but are disturbed by the inherent imprecision of the parameter estimates of the expo-power function being significantly different from zero.

7. Conclusion

Focusing on the preferences of private investors in stock markets is fundamental to any research in finance. In particular, we would like to understand why individuals act in the way they do and we would like to assist them by providing normative guidelines toward better (or optimal) behavior. Financial theory implicitly assumes utility maximization to obtain pricing kernels, from which the standard tools, such as stochastic discount factors, can be deployed to better understand how securities are priced and which class of investor assets are considered most valuable. To trace an arc from the literature on investor behavior in financial markets to methods developed and applied in experimental economics, we first presented a short summary on what we know about preferences and utility functions in financial markets and from where potential future streams of literature may spring. Then, we presented and adopted a popular and frequently applied econometric method from experimental economics, namely likelihood estimation and a famous application for model selection purposes. We check, based on a simulation study, whether a naive implementation of this method provides reliable results to address the research question, that is, to what extent it allows us to identify the correct utility model.

We find that for a very broad classification of utility models, this method provides acceptable outcomes. Yet, a closer look at the preference parameters reveals several caveats that come along with typical issues attached to financial data, and these issues may drive our results. In particular, deviations are attributable to effects stemming from multicollinearity and its coherent parameter identification problems, where some of these detrimental effects can be captured up to a certain degree by adjusting the error term specification. Furthermore, additional uncertainty stemming from changing market parameter estimates affects the precision of our estimates for risk preferences and cannot be simply remedied by using a higher standard deviation of the error term or a different assumption regarding

49 In this case, the solution would be acceptable if this estimator were not significantly different from zero, since it would imply that CRRA is the true model.
the stochastics of the error term. Particularly, if the variance of the error term becomes large, we detect a tendency to identify $SPT$ as utility model providing the best fit to data. We also find that a frequent issue, namely serial correlation of the residuals, does not seem to be significant. However, we detected a tendency to prefer nesting models over nested utility models, which is particularly prevalent if $RDU$ and $EXPO$ utility models are estimated along with $EUT$ and $CRRA$ utility models.
Appendix A. Utility Functions used in the Simulation Study

To substantiate the set of utility functions that specifies $\Delta(U_k|\theta_k)$, we consider several utility functions frequently used in the literature. For an Expected Utility-type investor (EUT), the preference over the risky outcomes of the stock are modeled as

$$U_{EUT}(W_t|\theta_{EUT}) = \sum_{j=1}^{t+1} \hat{p}_{j,t} u_{EUT}(W_t|\theta_{EUT}), \quad (A.1)$$

where $\hat{p}_{j,t}$ denotes the respective probabilities associated with the respective state.$^51$ We denote the utility functional as $u_{EUT}$ given the expo-power specification proposed by Saha (1993)$^51$

$$u_{EUT}(W_t|\theta_{EUT}) = 1 - e^{-\rho(W_t\hat{R}_{E,t}^{j+1}\hat{R}_D,t)^{1-\delta}} \rho^{-1}, \quad (A.2)$$

where $\rho$ governs relative and $\delta$ governs absolute risk aversion such that the term $u_{EUT}(W_t|\theta_{EUT})$ exhibits the properties of DARA for $1-\delta$ below unity and captures the behavior of CRRA if $1-\delta = 1$ and IARA for $1-\delta$ above 1. Regarding parameter $\rho$, functional (A.2) displays features of DRRA for $\rho < 0$ and IRRA if $\rho > 1$ (see also Saha et al. (1994), footnote 2). It is well established that $u_{EUT}(W_t|\theta_{EUT})$ converges to CRRA utility if $\rho$ approaches zero. To benchmark this case and to test for the hypothesis that the imprecision in estimating risk parameters is due to multicollinearity and its favoring of nesting utility, we explicitly model CRRA-utility, where we specify the utility functional as

$$u_{EUT}(W_t|\theta_{EUT}) = (W_t\hat{R}_{E,t}^{j+1}\hat{R}_D,t)^{1-\delta}(1-\delta)^{-1}, \quad (A.3)$$

Note that for $\delta = -1$, expression (A.3) covers mean-variance preferences (see for a proof Back (2012); for other characteristics regarding $\delta$ see Gollmer (2001)).

To capture the possible existence of generalized expected utility theories, we also consider an investor with Rank-dependent Utility (RDU) according to Quiggin (1993) and Wakker (1994), where the utility obtained from the risky asset is explicated as

$$U_{RDU}(W_t|\theta_{RDU}) = \sum_{j=1}^{t+1} \pi_{j,t}(\Delta \omega(\hat{p}_{j,t}|\theta_{RDU})) u_{RDU}(W_t|\theta_{RDU}). \quad (A.4)$$

To return from this generalized version to expected utility as presented above, we use the utility functionals as presented in equations (A.2) and (A.3) and define the decision weights $\pi_{j,t}(\omega(\hat{p}_{j,t}|\theta_{RDU}))$ as decumulative probability transformation functions according to Abdellaoui (2000) in the specification of the econometric model, to keep our results comparable to experimental evidence.$^{52}$ The probability

$^50$Technically, we specify the state probabilities as $\hat{p}_{j,t} = \left( (t+1) \right)^{t-j+1} \left( 1 - \hat{p}_t \right)^{-1}$. In an earlier version of our program, we calculated the respective values of $\hat{p}_{j,t}$ using Feller’s famous Reflection Principle (Feller (1968)) as $\hat{p}_{j,t} = \left[ \binom{t}{j} - \binom{t-1}{j-1} \right] \left( 1 - \hat{p}_t \right)^{-1}$. In testing our program, we find virtually no differences in the generated results between both specifications, such that we opted for the simpler binomial version of $\hat{p}_{j,t}$.

$^51$In the original paper, Saha (1993) suggests an exponential-power utility functional $u_{EUT}(W_t|\theta_{EUT}) = c - e^{-\rho(W_t\hat{R}_{E,t}^{j+1}\hat{R}_D,t)^{\delta}}$ where $\delta$ and $\rho$ are the respective parameters of this functional and $c$ denotes a constant. As Saha remarks (p. 906), setting $c = 1$ does not play a role in the characterization of risk attitudes or choices. In current research on asset pricing, this constant is usually set equal to zero and thus ignored.

$^52$More precisely, the decision weights are specified as $\pi_{j,t}(\Delta \omega(\hat{p}_{j,t}|\theta_{RDU})) = \omega \left( \sum_{k=j+1}^{t+1} \hat{p}_{k,t} \right) - \omega \left( \sum_{k=j+1}^{t+1} \hat{p}_{k,t} \right)$ for $j \leq t$ and for the highest node $\pi_{j,t}(\Delta \omega(\hat{p}_{j,t}|\theta_{RDU})) = \omega(\hat{p}_{t+1,t})$ for $j = t + 1$. In an earlier version of our program, we also implemented the logically equivalent
transformation function $\omega(\hat{p}_{j,t}(\theta_{RDU}))$, for which we identify two dominantly used versions from the experimental literature represents a central ingredient of RDU. Following the original article of Quiggin (1982), the decision weights are adapted from Karmarkar (1978) and Karmarkar (1979) and defined as

$$\omega(\hat{p}_{j,t}(\theta_{RDU})) = \hat{p}_{j,t}^\gamma (\hat{p}_{j,t}^\gamma + (1 - \hat{p}_{j,t})^\gamma)^{-1}. \quad (A.5)$$

It is noteworthy that the decision weights sum up to 1 and depend on the ranking of the states, even for $\gamma \neq 1$. Further, if $\gamma$ is equal to 1, RDU converges to EUT and the normal characterizations apply (Levy and Levy (2002a)). Studies on the impact of probability weighting, such as Barberis and Huang (2008) and Barberis (2012), instead impose a nonlinear weighting scheme, which can be reflected by the specification of $\pi_{j,t}(\omega(\hat{p}_{j,t}(\theta_{RDU})))$ as in Kahneman and Tversky (1979), explicitly stated in Tversky and Kahneman (1992), and used in Wu and Gonzalez (1996) whereas

$$\omega(\hat{p}_{j,t}(\theta_{RDU})) = \hat{p}_{j,t}^\gamma (\hat{p}_{j,t}^\gamma + (1 - \hat{p}_{j,t})^\gamma)^{-\frac{1}{\gamma}}, \quad (A.6)$$

in which the decision weights do not sum to unity. Note that the utility function $u_{EUT}(W_t|\theta_{RDU})$ remains the same as defined in EUT, since RDU reduces to expected utility if there is no probability weighting (e.g., if $\gamma = 1$ such that $\pi_{j,t}(\hat{p}_{j,t}(\theta_{RDU})) = \hat{p}_{j,t} \forall \hat{p}_{j,t} \in (0, 1)$). \(^{53}\)

As mentioned, some empirical studies on financial decision making suggest that there is some evidence that Prospect Theory may be at work in financial markets such that it seems advisable to model an investor’s preferences toward financial outcomes according to the original formulation of prospect theory, namely Simple Prospect Theory (SPT) as

$$U_{SPT}(W_t, W_{RP}|\theta_{SPT}) = \sum_{j=1}^{t+1}\pi_{j,t}(\omega(\hat{p}_{j,t}(\theta_{SPT})))u_{SPT}(W_t, W_{RP}|\theta_{SPT}), \quad (A.7)$$

where $W_{RP}$ marks a reference point (Kahneman and Tversky (1979)), assuming that preferences of the individual investor are based on changes of the initially invested wealth $W_0$ (for deviating reference points in stock markets such as historical extrema in prices Grinblatt and Keloharju (2001b), Garvey and Murphy (2004) or expectations Meng (2010)) and in which the decision weights $\pi_{j,t}(\omega(\hat{p}_{j,t}(\theta_{SPT})))$ are defined for each possible state (Kahneman and Tversky (1979)). \(^{54}\) In the original formulation of Kahneman and Tversky (1979) and as adapted in studies on various issues in finance (Berkelaar et al. (2004) with curvature parameter equal to 1, representation with a cumulative probability transformation function $\pi_{j,t}(\Delta\omega(\hat{p}_{j,t}(\theta_{RDU}))) = \omega\left(\sum_{k=1}^{t+1}\hat{p}_{k,t}\right) - \omega\left(\sum_{k=j}^{t+1}\hat{p}_{k,t}\right)$ for $j > 1$ and $\pi_{j,t}(\Delta\omega(\hat{p}_{j,t}(\theta_{RDU}))) = \omega(\hat{p}_{j,t})$ for $j = 1$ as in the original formulation of RDU according to Quiggin (1982) and Quiggin (1993). A test of our program reveals that the respective formulation of the decision weights appears to have no impact on our results, such that we refrain from an explicit distinction between these two possible cases.

\(^{53}\)By construction, under CRRA and RDU utility, the accrued return in the value function drops out in $\Delta_u(U_t|\theta_k)$. With regard to the aforementioned consistency of the estimation procedure, inclusion of $W_0$ in a linear-in-utility model does not invalidate the asymptotic consistency of the estimators $\hat{\theta}_{k,n,t}$. In highly nonlinear arguments, however, the consequences depend on whether there exists a value that causes the true model to coincide with any other utility model under consideration. In such case, including the level of wealth increases the sampling error of the model but does not affect its consistency (see Hensher and Johnson (1981)). Therefore, we add an increment of -0.0000001 to the cumulative returns in CRRA and RDU specifications. Based on a pretest, adding the increment has no significant effects on $\hat{\theta}_{k,n,t}$ but on the ranking according to the Akaike Criterion (AICC).

\(^{54}\)In detail, we establish a logical connection to RDU and define the decision weights as $\pi_{j,t} = \omega\left(\sum_{k=j+1}^{t+1}\hat{p}_{k,t}\right)$.
provided there exists a unique solution to the maximizing problem within the possible respective conditional probabilities as defined in (3.3). To clarify notation and functional values to particular values of \( t \) we use the same specification as for the original version of Prospect Theory. \( \gamma < CPT \), these decision weights sum to 1 if specified according to (A.5) and are usually \( k \) terms yielding the likelihood function for a utility model of type \( \theta \) investor or decision maker, given the assumption of stochastically independent error state-specific decision weights if the investor’s position in the risky asset generates specific decision weights in the domain of losses and as a cumulative function of the distinct from SPT, the difference of the probability weighting functions \( \omega \) with value functionals as defined in (A.8) and (A.9). Characteristic for CPT and \( \lambda \) loss aversion parameter, occur. Some financial studies, such as Barberis et al. (2001), Gomes (2005) and Barberis and Huang (2008) model the demand of SPT-type investors according to a different form of the value functional and apply a mathematical construct similar to CRRA utility, specified as

\[
u_{SPT}(W_t, W_{RP} | \theta_{SPT}) = \lambda I[\Delta W_t < 0] (|W_t \hat{R}^{t-j+1}_{D,t} \hat{R}^{-1}_{D,t} - W_{RP}|) ^{\alpha}, \tag{A.8}\]

where \( I[\Delta W_t < 0] \) represents an indicator taking the value of 1 if the change in wealth, measured as difference from the reference point \( \Delta W_t = W_t \hat{R}^{t-j+1}_{D,t} \hat{R}^{-1}_{D,t} - W_{RP} \), is negative and zero otherwise, indicating states where losses, weighted with loss aversion parameter \( \lambda \), occur.

It should be noted that, as in Barberis et al. (2001) and Barberis and Huang (2008), the individual’s consumption enters as arguments, although other variables and fundamentals may be considered (e.g., Barberis and Xiong (2009)). For Cumulative Prospect Theory (CPT), we calculate the utility of the financial prospects according to Tversky and Kahneman (1992) as

\[
u_{CPT}(W_t, W_{RP} | \theta_{CPT}) = \sum_{j=1}^{t+1} p_{j,t} (\hat{\omega}(\hat{p}_{j,t} | \theta_{CPT})) u_{CPT}(W_t, W_{RP} | \theta_{CPT}), \tag{A.10}\]

with value functionals as defined in (A.8) and (A.9). Characteristic for CPT and distinct from SPT, the difference of the probability weighting functions \( \hat{\omega}(\hat{p}_{j,t}) \) constitute rank-dependent decision weights as a decumulative function of the state-specific decision weights in the domain of losses and as a cumulative function of the state-specific decision weights if the investor’s position in the risky asset generates positive returns (Fennema and Wakker (1997)).\(^{55}\) It is noteworthy that for SPT and CPT, these decision weights sum to 1 if specified according to (A.5) and are usually subadditive under formulation (A.6) for \( \gamma < 1 \). Concerning \( u_{CPT}(W_t, W_{RP} | \theta_{CPT}) \), we use the same specification as for the original version of Prospect Theory.

Appendix B. Remarks on the Maximum Likelihood Approach

As elaborated, experimental studies maximize the overall likelihood of an investor or decision maker, given the assumption of stochastically independent error terms yielding the likelihood function for a utility model of type \( k \), expressed as

\[
\log L(\Delta_t(U_k | \theta_k)) = \sum_{t \in T} \sum_{i \in I_{k,t}} I_{k,t} \log p_{i_{k,t}}(\Delta_t(U_k | \theta_k)),
\]

in which it is required that \( \Delta_t(U_k | \theta_k) \) is a one-to-one relationship connecting the functional values to particular values of \( \theta_k \) and where \( p_{i_{k,t}}(\Delta_t(U_k | \theta_k)) \) denotes the respective conditional probabilities as defined in (3.3). To clarify notation and provided there exists a unique solution to the maximizing problem within the possible range of \( \theta_k \), maximizing the likelihood function (B.1) for a given sample and time

\(^{55}\)To be more precise, for CPT the decision weights are formulated as decumulative function \( \pi_{j,t} = \omega (\sum_{k=1}^{j} \hat{p}_{k,t}) - \omega (\sum_{k=1}^{j+1} \hat{p}_{k,t}) \forall j < [\hat{j}] \) and \( \pi_{j,t} = \omega (\hat{p}_{k,t}) \) if \( j = 1 \) for positive returns and specified as cumulative function \( \pi_{j,t} = \omega (\sum_{k=1}^{j+1} \hat{p}_{k,t}) - \omega (\sum_{k=1}^{j} \hat{p}_{k,t}) \forall j > [\hat{j}] \) and \( \pi_{j,t} = \omega (\hat{p}_{k,t+1}) \) if \( j = t+1 \) in the domain of losses where \( \hat{j} \) denotes the break-even node that classifies a state to be only in the realm of negative returns.
periods \( t \in \{1, \ldots, T\} \) returns a maximum likelihood estimate \( \hat{\theta}_{k|n,t} \), depending on the sample size, of the true but unknown parameter \( \hat{\theta}_k \), briefly denoted as

\[
\hat{\theta}_{k|n,t} = \arg \max_{\theta_k \in \theta_k} \log L(\Delta_t(U_k|\theta_k)). \tag{B.1}
\]

Accordingly, the obtained estimator \( \hat{\theta}_{k|n,t} \) is characterized by the usual standard conditions concerning the score vector \( S(\Delta_t(U_k|\theta_k)) \), which should be equal to a zero vector, and the Hessian matrix \( H(\Delta_t(U_k|\theta_k)) \), consequently being positive definite. Ignoring \( \sigma \) for a moment and following Edwards (1992), the score vector \( S(\Delta_t(U_k|\theta_k)) \) is

\[
S(\Delta_t(U_k|\theta_k)) = \sum_{I \in I_{k,t}} \delta(U_k|\theta_k) S(\Delta_t(\theta_k)) \tag{B.2}
\]

where we use the abbreviation \( \delta_t(U_k|\theta_k) \) to denote the square matrix of first derivatives of \( \Delta_t(U_k|\theta_k) \) with respect to each of its parameters and denote the \((K_k \times 1)\) vector of outer derivatives of the likelihood function as \( S(\Delta_t(\theta_k)) \), being the product of a diagonal matrix \( I \) with elements \( I_{k,t}/p_{k,t} \) and the diagonal matrix \( P_I \) containing the outer derivatives of \( p_{k,t} \). Following this notation, the Hessian matrix \( H(\Delta_t(U_k|\theta_k)) \) consists of two terms, namely a matrix containing partial derivatives of the elements of \( \delta(U_k|\theta_k) \) and a matrix collecting the second derivatives of \( \Delta_t(U_k|\theta_k) \) with respect to its parameters (see Edwards (1992) for details).\(^{56}\)

To obtain the Information matrix \( I(\Delta_t(U_k|\hat{\theta}_k)) \), the sign of the Hessian needs to be reversed and taken by its expectations, where we can use the fact that \( E(I_{k,t}) = p_{k,t} \). Since the sum of the choice probabilities equals 1 \( \sum_{t \in I_{k,t}} p_{k,t} = 1 \), the last term of the Hessian vanishes if evaluated at \( \hat{\theta}_k \) such that the last term can be greatly simplified (Fisher (1956), Edwards (1992), there Theorem 7.2.2) to

\[
I(\Delta_t(U_k|\hat{\theta}_k)) = \sum_{I \in I_{k,t}} \delta(U_k|\theta_k) I(\Delta_t(\theta_k)) \delta(U_k|\theta_k)'. \tag{B.3}
\]

Here, \( \delta_t(U_k|\theta_k) \) denotes the square matrix of first derivatives of \( \Delta_t(U_k|\theta_k) \) with respect to each of its parameters and \( I(\Delta_t(\theta_k)) = P_I P_I' I \) being the product of a diagonal matrix \( I \) with elements \( I_{k,t}/p_{k,t} \) and the diagonal matrix \( P_I \) containing the outer derivatives of \( p_{k,t} \). It is evident from this structure that for each \( I_{k,t} \)th term, the Hessian is a positive semi-definite matrix since \( I(\Delta_t(\theta_k)) = P_I P_I' I \) is symmetrical. Disregarding the possibility that \( H(\Delta_t(U_k|\theta_k)) \) is singular, the Hessian is in fact positive definite. This implies that \( I(\Delta_t(U_k|\theta_k)) \) is also a positive definite matrix over reasonable values of \( \hat{\theta}_k \).

### Appendix C. The Akaike Information Criterion and Model Tests used

According to the likelihood approach described above, it can be shown that a connection to the Akaike Information Criterion (Carbone and Hey (1994), Hey and Orme (1994), Carbone and Hey (1995) and Stott (2006)) can be established and used to identify the best-fitting utility function specification. To sketch this

\(^{56}\)Note that due to the independence assumption, each element of the score vector and the Hessian matrix consist of a series of sums. This is not surprising since, according to the independence assumption across time and choice sets, the log-likelihood function inherits the regularity property in the sense that differentiation and summation are interchangeable (e.g., Cramer (1986)), which in turn carries over to the entire sample if it holds for any single observation.
idea, recall that \( \log L(\Delta_t(U_k|\theta_k)) \) is continuous in \( \theta_k \) and twice differentiable.\(^{57} \) A second-order Taylor-expansion of the log-likelihood (B.1) around \( \hat{\theta}_k \) yields
\[
\log L(\Delta_t(U_k|\theta_k)) \approx \log L(\Delta_t(U_k|\hat{\theta}_k)) + S_{\hat{\theta}_k|n,t}(\Delta_t(U_k|\hat{\theta}_k)) + \frac{1}{2}nt(\Delta_t(U_k|\hat{\theta}_k) I(\Delta_t(U_k|\hat{\theta}_k))(\Delta_t(U_k|\hat{\theta}_k) - \Delta_t(U_k|\hat{\theta}_k))).
\]
(C.1)

If this expression is evaluated at \( \hat{\theta}_k \), the score vector \( S_{\hat{\theta}_k|n,t}(\Delta_t(U_k|\hat{\theta}_k)) \) equals zero and the Hessian can be rewritten as
\[
\log L(\Delta_t(U_k|\theta_k)) \approx \log L(\Delta_t(U_k|\hat{\theta}_k)) - \frac{1}{2}nt(\Delta_t(U_k|\hat{\theta}_k) I(\Delta_t(U_k|\hat{\theta}_k))(\Delta_t(U_k|\hat{\theta}_k) - \Delta_t(U_k|\hat{\theta}_k))).
\]

In this step, we make use of the fact that
\[
E\{nt(\Delta_t(U_k|\hat{\theta}_k|n,t) - \Delta_t(U_k|\hat{\theta}_k)) I(\Delta_t(U_k|\hat{\theta}_k|n,t))(\Delta_t(U_k|\hat{\theta}_k|n,t) - \Delta_t(U_k|\hat{\theta}_k))\}
\approx \text{tr}(J(\Delta_t(U_k|\hat{\theta}_k|n,t)) I(\Delta_t(U_k|\hat{\theta}_k|n,t))^{-1}),
\]
(C.2)
as shown elsewhere (Bozdogan (2000), Pawitan (2001)). \( J(\Delta_t(U_k|\hat{\theta}_k|n,t)) \) contains the product of the score vectors (Cramer (1986)) and can be written as \( J(\Delta_t(U_k|\hat{\theta}_k|n,t)) = E(S_{\hat{\theta}_k|n,t}(\Delta_t(U_k|\hat{\theta}_k)) S_{\hat{\theta}_k|n,t}(\Delta_t(U_k|\hat{\theta}_k))) \) and where \( tr \) denotes the trace of the product of the matrices within the brackets. Taking expectations, by (C.2), expression (C.1) can be rewritten as
\[
nt\hat{L}(\Delta_t(U_k|\theta_k)) \approx E(\log L(\Delta_t(U_k|\hat{\theta}_k))) - \frac{1}{2}nt \left( J(\Delta_t(U_k|\hat{\theta}_k|n,t)) I(\Delta_t(U_k|\hat{\theta}_k|n,t))^{-1} \right).
\]

If the number of observations or traded stocks increases beyond all bounds of the number of days \( t \) grows, according to Cramer (1986), \( J(\Delta_t(U_k|\hat{\theta}_k|n,t)) \approx I(\Delta_t(U_k|\hat{\theta}_k|n,t)), \) such that \( \text{tr} \left( I(\Delta_t(U_k|\hat{\theta}_k|n,t)) I(\Delta_t(U_k|\hat{\theta}_k|n,t))^{-1} \right) \), being the dimension of \( \hat{\theta}_k \), which is approximately equal to the number of parameters \( K_k \) of the respective utility model under consideration. Stated differently,
\[
\hat{L}(\Delta_t(U_k|\theta_k)) \approx -\frac{2\log L(\Delta_t(U_k|\hat{\theta}_k))}{nt} - \frac{2K_k}{nt}
\]
which is the information criterion according to Akaike (1974) in the representation of Amemiya (1980) as stated above, where we correct for the different number of observations by \( nt \). To contrast the results of the AIC by a finite correction version of Sugiura (1978) and Hurvich and Tsai (1989), we also invoke the corrected AIC, abbreviated as AICC, as
\[
\text{AICC} \equiv -\frac{2\log L(\Delta_t(U_k|\hat{\theta}_k))}{nt} + \frac{2K_k}{nt} + \frac{2K_k(K_k + 1)}{nt(n - K_k - 1)}.
\]
(C.3)

This form is usually recommended if the number of observations does not outweigh the number of parameters by more than a factor of 40 (Burnham and Anderson (2004)).

One serious drawback of the AIC, AICC, or basically any information criterion, is the fact that it, per se, cannot provide significance levels or statistical statements about how good the discrimination between the two competing models actually

\(^{57}\text{As for the derivatives of, for example, SPT towards some elements of } \theta_k, \text{ the derivative is not defined for some combinations of } R_{\theta_k}. \text{ Under a nonlinear probit model such as ours, the normal distribution is continuous, and given that summation and differentiation are interchangeable, the probability that those critical combinations appear converges to zero and thus can be ignored.} \)
is. To obtain the usual significance levels and to retrieve further information about the likelihood, the ranking according to the AICC is supplemented by the usual statistical tools made available by likelihood theory. For nested models where model \( k \) nests \( m \), the usual likelihood ratio test can be applied (Kent (1982)).

According to the null hypothesis \( H_0 \), both models are equally good in fitting the observed data such that the unconstrained maximum of \( \log L(\Delta_t(U_k(\theta_k))) \) should be close to the constraint maximum \( \log L(\Delta_t(U_m(\hat{\theta}_m))) \). The likelihood ratio test is calculated as

\[
LR_{nt}(\hat{\theta}_k, \hat{\theta}_m) = -2 \ln \left[ \frac{L(\Delta_t(U_k(\hat{\theta}_k)))}{L(\Delta_t(U_m(\hat{\theta}_m)))} \right]
\]

where under \( H_0 : LR_{nt}(\hat{\theta}_k, \hat{\theta}_m) \overset{L}{\to} \chi^2_{df_{k,m}} \)

for which it is well established that this ratio is non-negative and under \( H_0 \) asymptotically chi-square distributed with degrees of freedom \( df_{k,m} \) equal to the number of parameters of the unconstrained model minus the number of parameters of the constrained model (for a proof see, e.g., Rao (1973)). To derive p-values for contrasting non-nested models such as CRRA and CPT given \( W_0(RP) \), we apply a non-nested likelihood ratio test according to Vuong (1989), their Theorem 5.1., where we denote the maximized likelihood values of competing models \( k \) and \( m \), respectively. Vuong’s contribution is to show that under fairly general conditions and given that the null hypothesis holds, the expectations of the log-ratio of the two maximized likelihoods \( (\log L(\Delta_t(U_k(\hat{\theta}_k)))) \) and \( (\log L(\Delta_t(U_m(\hat{\theta}_m)))) \) for two competing models \( k \) and \( m \) should be zero. The expectations can be consistently estimated by the average of the likelihood ratio statistic over \( nt \) observations such that given the null hypothesis that the log of the likelihood ratio has an expectation of zero

\[
\frac{LR_{nt}(\hat{\theta}_k, \hat{\theta}_m)}{(\sqrt{nt})\hat{\omega}_{nt}} \overset{d}{\to} N(0, 1) \text{ with } \frac{1}{nt} LR_{nt}(\hat{\theta}_k, \hat{\theta}_m) \overset{L}{\to} E_0 \left[ \ln \frac{L(\Delta_t(U_k(\hat{\theta}_k)))}{L(\Delta_t(U_m(\hat{\theta}_m)))} \right].
\]

where

\[
\hat{\omega}_{nt}^2 = \frac{1}{n} \sum_{i=1}^{nt} \left[ \ln \left( \frac{L(\Delta_t(U_k(\hat{\theta}_k)))}{L(\Delta_t(U_m(\hat{\theta}_m)))} \right) \right]^2 - \frac{1}{nt} \sum_{i=1}^{nt} \ln \left( \frac{L(\Delta_t(U_k(\hat{\theta}_k)))}{L(\Delta_t(U_m(\hat{\theta}_m)))} \right)^2.
\]

for which it is shown that the resulting likelihood ratio statistic is asymptotically normally distributed. If the time series is long enough, the asymptotic properties might hold on the individual level as well.

---

58 Further, if two utility specifications, for instance model-\( k \) and competing model-\( m \), share the same number of risk preference parameters, then sorting them according to AICC leads to a selection of one model against the other. However, we find that, in some cases, where the highest AICC is not close to zero, the order in which these models are sorted can be due to a sufficiently small difference in the obtained log-likelihood, where the correction for the number of parameters strongly affects the ranking and overcompensates the difference in the log-likelihood. We check the ranking of utility models and use the Schwartz Information Criterion (also known as Bayes Information Criterion (Schwarz (1978))) as well as the original AIC. Since the penalty of a higher number of parameters according to the Schwartz Information Criterion is higher, we find that this effect is aggravated, though it affects the results only in exactly those cases pointed out above. In light of such deficiencies, to validate the obtained position, we supplement the results by appropriate significance tests between the first and second rank utility models to identify those models that are found to be hardly satisfactorily distinguished according to the AICC.

59 Technically, the Vuong test specifies that, under the null hypothesis, the expectation of the log ratio is symmetrically distributed around zero. In cases where the log of the likelihood ratio is not close to a normal distribution, alternative non-nested model tests have been proposed (e.g., Clarke (2003) and Clarke (2007)).

60 The fact that we use time-series observations does not invalidate the possibility of applying the AIC, since Akaike (1978) provides a time-series version of the information criterion.
REFERENCES


Experimental econometrics for finance - Analysis of a Likelihood Approach


Mann, H. and D. Whitney (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* 18, 50–60.


Economic Review 98(1), 38–71.


Recent Issues

No. 147  Andreas Hackethal, Sven-Thorsten Jakusch, Steffen Meyer
Taming all Investors with the same Brush? Evidence for Heterogeneity in Individual Preferences from a Maximum Likelihood Approach

No. 146  Sven-Thorsten Jakusch, Steffen Meyer, Andreas Hackethal
Taming Models of Prospect Theory in the Wild? Estimation of Vlcek and Hens (2011)

No. 145  Christian Geppert, Alexander Ludwig, Raphael Abiry
Secular Stagnation? Growth, Asset Returns and Welfare in the Next Decades: First Results

No. 144  Mario Bellia, Loriana Pelizzon, Marti G. Subrahmanyam, Jun Uno, Darya Yuferova
Low-Latency Trading and Price Discovery: Evidence from the Tokyo Stock Exchange in the Pre-Opening and Opening Periods

No. 143  Peter Gomber, Satchit Sagade, Erik Theissen, Moritz Christian Weber, Christian Westheide
Spoilt for Choice: Order Routing Decisions in Fragmented Equity Markets

No. 142  Nathanael Vellekoop
The Impact of Long-Run Macroeconomic Experiences on Personality

No. 141  Brigitte Haar
Freedom of Contract and Financial Stability through the lens of the Legal Theory of Finance

No. 140  Reint Gropp, Rasa Karapandza, Julian Opferkuch
The Forward-Looking Disclosures of Corporate Managers: Theory and Evidence

No. 139  Holger Kraft, Claus Munk, Farina Weiss
Predictors and Portfolios Over the Life Cycle

No. 138  Mohammed Aldegwy, Matthias Thiemann
How Economics Got it Wrong: Formalism, Equilibrium Modelling and Pseudo-Optimization in Banking Regulatory Studies

No. 137  Elia Berdin, Cosimo Pancaro, Christoffer Kok
A Stochastic Forward-Looking Model to Assess the Profitability and Solvency of European Insurers

No. 136  Matthias Thiemann, Mohammed Aldegwy, Edin Ibrocevic
Understanding the Shift from Micro to Macro-Prudential Thinking: A Discursive Network Analysis

No. 135  Douglas Cumming, Jochen Werth, Yelin Zhang
Governance in Entrepreneurial Ecosystems: Venture Capitalists vs. Technology Parks