Appendix B: Overview of Horse Race Studies

1. Studies focusing on different stochastic specifications

Blavatskyy (2013) compares eight deterministic models and a simple heuristic. He uses the stronger utility specification for the stochastic component. Results suggest that EUT, RDU and the heuristic achieve the best fits: together they can explain around a quarter of observed behavior. The prominent losers are EV, DA and PR. Moreover, QU and WU are only the best models for one or two subjects. Overall, the analysis suggests that EV, DAT, PR, QU and WU can be discarded from the menu of promising descriptive theories. Blavatsky (2007) conducts a parametric estimation of his specification (StEUT) and compare it only with CPT (identical to RDUT if the experiment includes lotteries with nonnegative outcomes) using ten datasets from Conlisk (1989), Kagel *et al* (1990), Camerer (1989, 1992), Tversky and Kahneman (1992), Camerer and Ho (1994), Hey and Orme (1994), Wu and Gonzalez (1996), Loomes and Sugden (1998) and Gonzalez and Wu (1999). Overall, according to Vuong's likelihood ratio test the descriptive performance of both specifications is equally good. The Schwarz criterion favours StEUT while the Akaike criterion favours CPT. Blavatskyy (2018) uses stronger utility to evaluate and compare his new model, that is, second-generation disappointment aversion theory (DAT2), with other models in the literature. The new model is reminiscent of a model presented in Hagen (1991) and performs better than others. Finally, Hey *et al* (2010), focusses on decision under ambiguity and compares the models using contextual and strong utility stochastic specifications.

2. Studies focusing on a richer menu of stochastic specifications

Wilcox (2008, 2011) compares his contextual utility specification with the strong/strict utility specification and the wandering vector specification. His study departs from other studies by not merely focusing on in-sample but also out-of-sample log-likelihoods. The former is the common practice in the literature and is done by comparing log-likelihoods resulting from parameters estimated by maximum likelihood using the whole data, that is, it focusses on the *descriptive pow*er of specifications. Out-of-sample fit evaluates the *predictive power* of specifications by comparing predictive log-likelihoods. To calculate predictive log-likelihoods, a subset of data is used to estimate the parameters of models. Then these estimates are used to calculate the log-likelihoods on the remaining portion of the data to construct test statistics based on the likelihood values. Overall, contextual utility performs better in both categories whether it is combined with EUT or RDUT. When we look at only in-sample fit, strict utility exhibits the worst fit for both deterministic models. For out-of-sample fit, it is the worst performer for RDUT as well, but, for EUT, the wandering vector is the poorest.

Blavatsky and Pogrebna (2010) focus on seven deterministic models and combines them with five stochastic specifications (tremble, strong utility with homoscedastic, heteroscedastic and heteroscedastic-truncated random errors and RPM). They estimated parametric forms of EV, EUT with constant relative risk aversion (CRRA), and expo-power (EP) utility functions, RT or skew-symmetric bilinear utility theory (SSB, hereafter) (Fishburn, 1983), YDM, RDUT and DAT. Their data is distinct in the sense that it is not from a laboratory experiment, instead they use data from a natural experiment: 114 (Italian version) and 518 (British version) episodes of a TV show called *"Deal or No Deal".*

The data considers decisions when choosing between a risky lottery with large stakes and a certain amount of money. Monetary outcomes vary between 1 cent and half-a-million euros hidden in separate but identical boxes. At the beginning of the game, each contestant is endowed with a box and at each stage, he or she forgoes a box by opening and revealing the prize in it. After each opening, the contestant receives an offer of sure amount of money. He or she can either accept the offer and leave or continue opening the boxes. The game ends when all the boxes are opened or the contestant has accepted an offer.

The analysis highlights the importance of the stochastic specification: for example, different stochastic specifications lead to different estimates of the risk attitude. For example, EUT with a CARA utility function exhibits risk-neutrality with a tremble specification, risk-aversion with the strong utility specification and risk-seeking with the random utility specification. As far as goodness-of-fit is concerned, the winners are the strong utility specification with heteroscedastic truncated errors and RPM.

Blavatsky (2011) compares his Model 1 with strong utility, contextual utility and IEUA, using EUT with a Bernoulli utility function. He finds that his Model 1 outperforms the other specifications. Contextual utility and IEUA specifications have a similar performance, whereas all specifications outperform the strong utility specification.

Blavatsky (2014) compares the goodness-of-fit of five stochastic specifications using EUT and RDUT on Hey and Orme's (1994) dataset. These stochastic specifications are strong, stronger, contextual utility and the incremental EU advantage specification (Fishburn, 1978) and "Model 1" in Blavatsky (2009, 2011).

In the case of EUT, results suggest that the stronger utility specification performs better than the strong and contextual utility specifications. The stronger utility specification has a similar goodness-of-fit to that of Fishburn's specification when the Akaike information criterion is used, but is better when the number of parameters is penalised more, for example, using the Schwarz information criterion. When the stronger utility specification is compared to Model 1, there seems to be no improvement in goodness-of-fit under EUT. Compared to Model 1, the stronger utility specification can only accommodate violations of expected utility when the preference function is nonlinear in probabilities. However, when we use RDUT with the Quiggin weighting function, the stronger utility specification achieves a slightly better performance than Model 1. Similar results are found when EUT is the deterministic structure.

Wilcox (2015) compares strong utility, contextual utility, DFT and stronger utility, within EUT and RDUT using a new dataset composed of 80 subjects' responses for 100 pairs of lotteries. Ranking among the stochastic specifications depends on the way we specify the functional form of a deterministic model. For example, when EU and RDUT has a parametric form, CU performs better than DFT and stronger utility, whereas stronger utility achieves the best fit under non-parametric estimation for EUT while DFT is the best for RDUT.

Table B1: Datasets			
Acr	**Source**	**References**	**Notes**
HO	Hey & Orme (1994)	Hey (1995); Carbone & Hey (2000); Buschena & Zilberman (2000); Wilcox (2008); Wilcox (2011); Blavatsky (2011); Blavatsky (2014)	80 subjects participated in 4 experiments in different occasions: *Circles 1*, *Dynamics 1*, *Circles 2*, *Dynamics* 2. Only *Circles 1* and *2* used in analysis, which composed of same 100 pairwise choice questions in different order.
23D	Harless & Camerer (1994)	-	23 datasets aggregated from previous studies consisting of 2000 choice patterns
CH	Carbone & Hey (1995)	Carbone (1997)	40 subjects answered 94 pairwise choice questions; in 8 of them one lottery dominates other.
LS	Loomes & Sugden (1998)	Loomes *et al* (2002)	92 subjects answered 45 pairwise choice questions twice.
10D	Blavatskyy (2007)	-	10 datasets from previous studies eliciting certainty equivalents, binary choices in hypothetical or incentivised settings, so each dataset analyzed separately.
TV	Blavatskyy & Pogrebna (2010)	-	Consists of 114 and 518 episodes of TV show called "Deal or No Deal"; aired in Italy and UK, respectively.
H10	Hey *et al* (2010)	-	Unlike others it is an experiment on ambiguity; 48 subjects answered 162 pairwise choice questions; 3 treatments vary according to the total number of balls.
B13	Blavatskyy (2013)	Blavatskyy (2018)	38 subjects answered 140 binary choice questions
W15	Wilcox (2015)		80 subjects answered 100 binary choice questions

Notes: First column lists the acronyms of datasets, second column lists the study in which the dataset is first used, third column lists the studies that use the dataset. Final column includes some basic information about each dataset.

Table B2: Snapshot of comparative studies

Study	Year	Deterministic Models	Tre	HoS	HeS	THeS	RPM	Str	WV	CU	DFT	M1	Ser	Data
Hey & Orme	1994	EV, EUT, DAT, PR, QU, RT, RDUT, WU, YDM		✓										O
Harless & Camerer	1994	EV, EUT, fanning in and out, RDUT, PT	✓											23D
Hey	1995	EUT , EV, DAT, PR, QU, RT, RDUT, WU		✓	✓									HO
Carbone	1997	EUT	✓	✓			✓							CH
Carbone & Hey	2000	EV, EUT, DAT, PR, RT, RDUT, QU, WU	✓	✓										HO
Buschena & Zilberman	2000	EUT, DAT, PR, RT, RDUT, QU, WU	✓	✓										HO
Loomes *et al*	2002	EUT, RDUT	✓	✓			✓							LS
Blavatsky	2007	StEUT, CPT		✓										10D
Wilcox	2008	EUT, RDUT		✓			✓	✓	✓	✓				HO
Blavatsky & Pogrebna	2010	EV,EUT, RT YDM, RDUT, DAT	✓	✓	✓	✓	✓							TV
Hey *et al*	2010	EV, EUT, CEU, PT, CPT, DFT, GSMaxMin, GSMaxMax, Alpha, MinReg, MaxMax, MaxMin		✓						✓				O
Wilcox	2011	EUT, RDUT		✓			✓	✓		✓				HO
Blavatsky	2011	EUT		✓						✓		✓		HO
Blavatsky	2013	EV, EUT, DAT, PR, QU, RDUT, WU, YDM, MVA, H											✓	O
Blavatsky	2014	EUT, RDUT		✓						✓		✓	✓	HO
Wilcox	2015	EUT, RDUT		✓						✓	✓		✓	O
Blavatsky	2018	DAT2, EV, EUT, DAT, PR, QU, RDUT, WU, YDM, H											✓	B13

Column 3 EV: *expected value*; EUT: *expected utility*; DAT: *disappointment aversion*; PR: *prospective reference*; QU: *quadratic utility*; RT: *regret theory*; RDUT: *rank dependent utility*; WU: *weighted utility*; YDM: *Yaari's dual model*; PT: *prospect theory*; StEUT: *stochastic expected utility*; CPT: *cumulative prospect theory*; GSMaxMin: *Gilbao and Schmeidler's MaxMin* ; GSMaxMax: *Gilbao and Schmeidler's MaxMax*; Alpha: *Ghirdato et al's alpha*; MinReg: *minimum regret*; MVA: *mean variance approach*; H: *a heuristic*: DAT2: *second generation disappointment aversion theory.* **Coloumns 4 to 14** Tre: *Tremble*; HoS: *homoscedastic strong utility*, HeS: *heteroscedastic strong utility*, THeS: *truncated and heteroscedastic strong utility*; RPM: *random preference model*; Str: *strict utility*; WV: *wandering vector*; CU: *contextual utility*; DFT: *decision field theory*; M1: *model 1 of Blavatsky*; Ser: *stronger utility model*. **Coloumn 15** See Table B1 for acronym

References for Appendix B

Blavatskyy, P. R. (2007). Stochastic expected utility theory. Journal of Risk and Uncertainty, 34(3), 259-286.

Blavatskyy, P. R. (2008). "Probabilistic choice and stochastic dominance," IEW Working Paper 364, available at http://www.iew.uzh.ch/wp/iewwp364.pdf.

Blavatskyy, P. (2009). Preference reversals and probabilistic decisions. Journal of Risk and Uncertainty, 39(3), 237–250.

Blavatskyy, P. R., & Pogrebna, G. (2010). Models of stochastic choice and decision theories: Why both are important for analyzing decisions. Journal of Applied Econometrics, 25(6), 963-986.

Blavatskyy, P. (2011). A model of probabilistic choice satisfying first-order stochastic dominance. Management Science, 57(3), 542–548.

Blavatskyy, P. R. (2012). Probabilistic choice and stochastic dominance. Economic Theory, 50(1), 59-83.

Blavatskyy, P. (2013). Which decision theory?. Economics Letters, 120(1), 40-44.

Blavatskyy, P. R. (2014). Stronger utility. Theory and Decision, 76(2), 265-286.

Blavatskyy, P. (2018). A second-generation disappointment aversion theory of decision making under risk. Theory and Decision, 84(1), 29-60.

Camerer, C. F. (1989). An experimental test of several generalized utility theories. Journal of Risk and uncertainty, 2(1), 61-104.

Camerer, Colin F. (1992). "Recent Tests of Generalizations of Expected Utility Theory." In Ward Edwards (ed.), Utility: Theories, Measurement, and Applications. Norwell, MA: Kluwer.

Camerer, Colin F., and Teck-Hua Ho. (1994). "Violations of the Betweenness Axiom and Nonlinearity in Probability," Journal of Risk and Uncertainty 8, 167–196.

Carbone, E., Hey, J.D. (1995). A comparison of the estimates of EU and non-EU preference functionals using data from pairwise choice and complete ranking experiments. Geneva Papers on Risk and Insurance Theory 20, 111–133.

Carbone, E. (1997). Investigation of stochastic preference theory using experimental data. Economics Letters, 57(3), 305-311.

Carbone, E., & Hey, J. D. (2000). Which error story is best? Journal of Risk and Uncertainty, 20(2), 161-176.

Conlisk, John. (1989). "Three Variants on the Allais Example," American Economic Review 79, 392–407.

Fechner, G. (1966). Elemente der Psychophysik. Leipzig: Breitkopf and Hartel, 1860. Translated by Helmut E. Adler as Elements of Psychophysics (New York: Holt, Reinhart and Winston).

Fishburn, P. (1978). A probabilistic expected utility theory of risky binary choices. International Economic Review, 19, 633–646.

Fishburn P. (1983). Transitive measurable utility. Journal of Economic Theory 31: 293–317.

Gonzalez, Richard, and George Wu. (1999). "On the Shape of the Probability Weighting Function," Cognitive Psychology 38, 129–166.

Gul, F., & Pesendorfer, W. (2006). Random expected utility. *Econometrica*, *74*(1), 121-146.

Hagen, O. (1991). "Decisions under risk: A descriptive model and a technique for decision making," European Journal of Political Economy, 7(3), 381-405.

Harless, D. W., & Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. Econometrica: Journal of the Econometric Society, 1251-1289.

Hey, J. D., & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. Econometrica: Journal of the Econometric Society, 1291-1326.

Hey, J. D. (1995). Experimental investigations of errors in decision making under risk. European Economic Review, 39(3-4), 633-640.

Hey, J. D., Lotito, G., & Maffioletti, A. (2010). The descriptive and predictive adequacy of theories of decision making under uncertainty/ambiguity. *Journal of risk and uncertainty*, *41*(2), 81-111.

Kagel, John H., Don N. MacDonald, and Raymond C. Battalio. (1990). "Tests of 'Fanning Out' of Indifference Curves: Results from Animal and Human Experiments," American Economic Review 80, 912–921.

Loomes, G., & Sugden, R. (1995). Incorporating a stochastic element into decision theories. European Economic Review, 39(3), 641-648.

Loomes, G., & Sugden, R. (1998). Testing different stochastic specifications of risky choice. Economica, 65(260), 581-598.

Loomes, G., Moffatt, P. G., & Sugden, R. (2002). A microeconometric test of alternative stochastic theories of risky choice. Journal of risk and Uncertainty, 24(2), 103-130.

Loomes, G., Rodríguez-Puerta, I., & Pinto-Prades, J. L. (2013). Comment on "A model of probabilistic choice satisfying first-order stochastic dominance" by Pavlo Blavatskyy. Management Science, 60(5), 1346-1350.

Melkonyan, T., & Safra, Z. (2015). Intrinsic variability in group and individual decision making. Management Science, 62(9), 2651-2667.

Tversky, Amos, and Daniel Kahneman. (1992). "Advances in Prospect Theory: Cumulative Representation of Uncertainty," Journal of Risk and Uncertainty 5, 297–323.

Wilcox, N. T. (2008). Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison. In Risk aversion in experiments (pp. 197-292). Emerald Group Publishing Limited.

Wilcox, N. T. (2011). ‘Stochastically more risk averse: A contextual theory of stochastic discrete choice under risk. Journal of Econometrics, 162(1), 89-104.

Wilcox, N. T. (2015). 'Error and generalization in discrete choice under risk'. Working paper. Economic Science Institute, Chapman University.

Wu, George, and Richard Gonzalez. (1996). “Curvature of the Probability Weighting Function,” Management Science 42, 1676–1690.