WARWICK

ECONOMICS

CRETA

Centre for Research in Economic Theory and its Applications

Discussion Paper Series

Generalizing Heterogeneous Dynamic Heuristic Selection

Giorgos Galanis, Iraklis Kollias, Ioanis Leventidis & Joep Lustenhouwer

August 2022 No: 73

CRETA
Centre for Research in Economic Theory and its Applications
Department of Economics
University of Warwick, Coventry,
CV4 7AL, United Kingdom

warwick.ac.uk/fac/soc/economics/research/centres/creta/papers

Generalizing Heterogeneous Dynamic Heuristic Selection*

Giorgos Galanis[†], Iraklis Kollias[‡], Ioanis Leventidis[§], Joep Lustenhouwer[¶]

Abstract

The growing literature in behavioral finance and macroeconomics that uses dynamic discrete choice models has overwhelmingly assumed that individual choices are made on the basis of a logit framework. While this assumption allows for analytical tractability, it comes with a number of restrictions with regards to the economic environments it can represent. These restrictions are lifted if a probit framework is used instead. In this paper we compare the two approaches and show that, due to its ability to allow for correlations between the random part of different choice alternatives as well as random taste variation, the probit-based model can better fit actual choice data from an existing laboratory experiment, especially if there are more choice alternatives. On the other hand, for the case of two choice alternatives without random taste variation, the probit-based and logit-based models result in very similar dynamics. But even in that case, we find that important qualitative differences arise – in terms of an additional region of chaos – in the cobweb model of the seminal work of Brock and Hommes (1997). Our work highlights the usefulness of using the probit framework for extensions of existing theoretical models and as a way to better fit dynamic experimental or real world choice data.

*We are grateful to Mikhail Anufriev, Kenneth Train, Cars Hommes, Giorgio Ricchiuti, Kostas Zachariadis as well as the participants of the 25th CEF conference (virtually) and the 24.5th Workshop on Economics with Heterogeneous Interacting Agents (virtually) for very helpful comments. We further thank Mikhail Anufriev, Te Bao and Jan Tuinstra for sharing their experimental data and estimation codes.

[†]School of Business and Management, Queen Mary, University of London; Centre for Research in Economic Theory and its Applications, University of Warwick, UK

[‡]Department of Economics, University of Athens, Greece

[§]Department of Economics, University of Athens, Greece

[¶]Department of Economics, Heidelberg University, Germany

1 Introduction

Over the last decades, large literatures on behavioral finance and behavioral economics have been developed that aim to provide more robust modeling and policy implications by taking account of deviations of full rationality of financial and economic actors (see e.g. De Bondt and Thaler, 1985; De Long et al., 1990; Shleifer and Vishny, 1997; Bordalo et al., 2018; Evans and Honkapohja, 2012; Woodford, 2013; Azeredo da Silveira and Woodford, 2019; and Gabaix, 2020). A growing literature within behavioral finance and behavioral macroeconomics, starting with Brock and Hommes (1997, 1998), has incorporated behavioral insights like bounded rationality (Simon, 1957, 1979) and the use of heuristics (Kahneman and Tversky, 1973; Kahneman, 2003) into the modeling of expectation formation. Models of this type rely on a discrete choice framework (Manski and McFadden, 1981; McFadden, 2001) and assume that individuals have heterogeneous expectations and choose from a set of prediction rules. The fractions of individuals that follow the different prediction rules in a given period directly influences aggregate forecasts and so determines the values of the financial or macroeconomic variables in the model. Crucially, the choice of prediction rules is based on their relative performance (fitness) through a discrete choice model. This setting, where the choice of prediction rules determines model outcomes and where these model outcomes subsequently impact on future prediction rule choices results in a dynamic switching framework. This framework has not only allowed for more realistic assumptions regarding prediction rules observed in experiments (for example see Hommes et al., 2005; Anufriev et al., 2016, 2018) but also is able to capture several empirically observed phenomena (for example see Branch, 2004; Hommes, 2006; Heemeijer et al., 2009; Lux, 2009; Cornea-Madeira et al., 2019; Hommes, 2021).

The standard assumption within this literature is that switching between different expectation rules takes the functional form that corresponds to choices being made on the basis of a logit discrete choice model. This (implicit) assumption is convenient as it allows for closed form expressions of the choice probabilities; however, it comes with some limitations. In particular, the logit framework assumes that the unobservable factors

which influence choices between alternatives are independent and identically distributed (Train, 2009). As a consequence, random variation across tastes is not allowed for. Moreover, the logit framework is restricted to feature the independence of irrelevant alternatives property, which can be a restrictive assumption. Furthermore, Anufriev et al. (2016) show that the standard logit model is not able to capture well the switching behavior when more choice alternatives are available.

In this paper we use, instead, a probit framework which does not suffer from these limitations. We show that this framework is better able to capture switching behavior documented in the labratory experiment of Anufriev et al. (2016) for more than two choice alternatives. We provide intuition for this result by investigating in detail (i) how the probit model allows for deviations from the independence of irrelevant alternatives property and (ii) what the implications of allowing for random taste variation are. Note that the above limitations can also be overcome by using mixed logit, instead. Our choice of probit is due to the fact that it is a widely used fully specified model, whereas opting for mixed logit would require additional assumptions regarding the distribution of the random coefficients. In the rest of this paper, when we refer to the logit model, we have in mind the standard logit model without the extensions of mixed-logit.

We further show that for the case of two choice alternatives and no random taste variation, the probit- and logit-based heuristic switching models are very similar. As a consequence, the key insights from the seminal paper of Brock and Hommes (1997), which shows the emergence of chaotic price dynamics, carry over to the case of a probit-based heuristic switching mechanism. However, even in that case, important qualitative differences may arise depending on whether the logit- or probit-framework is chosen. In particular, we show that, under the probit-based framework, an additional region of chaos arises in a relevant region of the parameter space. The key implication of our results is that the relevant literature should be using a probit framework at least a robustness check especially in cases where the emergence of chaos is discussed and in situations where more than two alternatives are available.

Our results, showing the usefulness of a probit framework for the experiment of

Anufriev et al. (2016), contribute to the behavioural/experimental literature on asset pricing with agents choosing among heterogeneous expectation rules (Hommes et al., 2005; Anufriev and Hommes, 2012; Bao et al., 2012; Anufriev et al., 2018) and where the switching process assumes a logit framework. In this way, our work is connected to the broader experimental literature on learning to forecast (Colasante et al., 2017; Bao et al., 2017; Kopányi-Peuker and Weber, 2020). By showing the similarities between the logit and the key insights of Brock and Hommes (1997, 1998) hold even under more general economic assumptions and (ii) provide a robustness criterion for models in this tradition which have been used to study a number of economic and social phenomena. Examples of applications of dynamic discrete choice models include, social interactions (Brock and Durlauf, 2001) asset pricing (Chiarella and He, 2002, 2003), the effectiveness of financial transaction taxes (Westerhoff and Dieci, 2006), macroeconomic activity (De Grauwe, 2012), monetary policy (De Grauwe, 2011; Hommes and Lustenhouwer, 2019; Hommes et al., 2019; Assenza et al., 2021), exchange rate dynamics (De Grauwe and Grimaldi, 2005), fiscal policy (Hommes et al., 2018), real-financial interactions (Flaschel et al., 2018), voting (Di Guilmi and Galanis, 2021) and physical distancing in response to COVID-19 (Di Guilmi et al., 2020).[1]

The structure of the rest of the paper is as follows. The next section introduces the behavioral microfoundations of discrete choice models and discusses what the economic assumptions are upon which logit and probit models are based. In Section 3, we compare the logit- and probit-frameworks, especially focusing on the independence of irrelevant alternatives, as well as their fit to the Anufriev et al. (2016) experimental data. Section 4 focuses on random taste variation in a probit-framework. In Section 5, we present a probit version of the Brock and Hommes (1997) model and analyze its dynamic equilibrium properties. The final section concludes.

[1] As the relevant literature is too extensive to review, we refer the reader to the reviews of Hommes (2006); Chiarella et al. (2009); Branch and McGough (2018); Dieci and He (2018) for models in finance and macroeconomics.

2 Random Utility Models

Assume an environment with a large number N of individuals who face a choice between mutually exclusive alternatives from a set J. An individual $i \in \{1, \ldots, N\}$ would choose alternative A in period t iff

$$U_t^{Ai} > U_t^{ki}, \qquad k \in J \ with \ k \neq A, \tag{1}$$

where U_t^{ji} is the utility of individual i from choosing $j \in J$ in t, and is given by

$$U_t^{ji} = \beta_i x_{j,t} + \epsilon_t^{ji}, \tag{2}$$

where $\beta_i x_{j,t}$ captures the observable attributes that affect i's decision and ϵ_t^{ji} the unobservable ones which are assumed to vary across individuals. $x_{j,t}$ is a column vector of observable variables at t which affect the decision between the different alternatives and β_i is a row vector which captures the importance and type of effect (positive or negative) of the each of the variables in $x_{j,t}$. For example, in Flaschel et al. (2018) the decision of agents between adopting a fundamentalist or chartist rule depends on relative economic performance, profitability, price volatility and the shares of investors that follow the different strategies.

When decisions are made based on a weighted average of past forecast errors, the first element of $x_{j,t}$ in (2) corresponds to the most recent forecast error, whereas the other element(s) of $x_{j,t}$ capture forecast errors further in the past. Correspondingly, the first element of β_i is the weight that is put on the most recent forecast (e.g. $(1-\rho)$ with ρ a memory parameter). Other examples include the case where forecasters forecast multiple variables and switch their forecasting strategy jointly on all these variables, as in e.g. Hommes and Lustenhouwer (2019). Forecast errors of different variables then can be seen as taking different entries in the vector $x_{j,t}$ with corresponding weights in the vector β_i.

Note that we can identify two types of heterogeneity across individuals related to

preferences with respect to both observable factors through β_i and unobservable ones through ϵ_t^{ji}. The first type of heterogeneity could be related both to different preferences (for example different levels of risk aversion) and to behavioral factors related to the importance individuals may give to the various variables of $x_{j,t}$. The second type of heterogeneity captures different types of individual biases which are not related to the observable factors and vary over the population.

In order to calculate the probability that a particular agent chooses a certain choice alternative it is necessary to make some concrete assumptions regarding ϵ_t^{ji}. We will next discuss the specific assumptions in the most standard discrete choice models: logit and probit.

2.1 Logit

The logit model is based on the assumption that ϵ_t^{ji} are independent and identically distributed (IID) and follow an extreme value distribution, which has the property that the difference between ϵ_t^{Ai} and ϵ_t^{Bi}, i.e., $\epsilon_t^{ABi} = \epsilon_t^{Ai} - \epsilon_t^{Bi}$ follows the logistic distribution for all t. As it is obvious from (1), it is only the *relative* value of the utilities that influence individual decisions, hence this modeling approach to discrete choices has important advantages due to the fact that it allows for analytical solutions as we discuss in more detail below. However, the analytical tractability comes with restrictions regarding the economic environments that it can represent.

The main reason for this is the IID assumption of the error terms that must be imposed. This puts limits on the exact form of heterogeneity between individuals that may arise. More concretely, in the applications that we are interested in, it must be assumed that the heterogeneity between individuals materializes only in ϵ_t^{ji}, whereas $\beta_i = \beta$ must be equal for all individuals.[2] The reason for this can be seen from rewriting

[2]More generally, heterogeneity in β_i is only allowed if it is directly linked to observable characteristics of the individual agents populating the model. However, in the way that the discrete choice model is used in behavioral macroeconomics and finance, no explicit assumptions are made about the source of heterogeneity across agents or their individual characteristics. Hence, the option of having heterogeneity in β_i that is directly related to *observable* characteristics is eliminated and it is assumed that there is a β that is common across agents. See Train (2009) for details.

the utility of individual i from choice j at time t as

$$U_t^{ji} = \bar{\beta} x_{j,t} + \tilde{\beta}_i x_{j,t} + \epsilon_t^{ji}, \tag{3}$$

where $\tilde{\beta}_i = \beta_i - \bar{\beta}$. Let $\eta_t^{ji} = \tilde{\beta}_i x_{j,t} + \epsilon_t^{ji}$, so that the utility can be expressed as

$$U_t^{ji} = \bar{\beta} x_{j,t} + \eta_t^{ji}. \tag{4}$$

In principle, this would be a form that could be used for estimating the logit model. However, given that $\tilde{\beta}_i$ is now implicitly part of the error term η_t^{ji}, these error terms cannot possibly be IID distributed anymore. Hence, for logit, any specification where individual characteristics become part of the error terms leads to a misspecification of the model.[3] As a consequence, the logit framework does not allow for variation in tastes.

In terms of the heuristic switching models used in behavioral macroeconomics and finance, it is far from clear that all the heterogeneity amongst people that we are interested in can be captured by the additive term ϵ_t^{ji}. In the real world, heterogeneity between investors and economic participants would also take the form of different weights on the different components of $x_{j,t}$. For example, there may be heterogeneity on the rate of discounting of past forecast errors, or investors may have different risk aversion coefficients.

Finally, the logit model implicitly assumes *independence of irrelevant alternatives*, or IIA. In some situations this may be a realistic assumption, but this is not always the case. We can think of the following expectation formation parallel to McFadden's (1974) Red Bus/ Blue Bus example. Assume that investors face a choice between costly rational expectation and a non-costly adaptive expectation with a weight $\mu_1 \in (0, 1)$ capturing the adjustment process. The IIA would imply that if another type of expectation, for example an adaptive one with parameter $\mu_2 \neq \mu_1$, was an available option, this would

[3]Another limitation of the logit model that stems from the IID assumption is that the ϵ_t^{ji} terms may not be assumed to be autocorrelated. This assumption may be especially limiting in dynamic models where only a small set of observable variables is assumed. Here, the unobservable component may refer to behavioral biases which remain the same over time. Individuals may have a very strong preference towards adopting one strategy independent of the observable characteristics. For example, in a framework like Brock and Hommes (1997) where it is costly to be 'fundamentalist', a subset of investors may never choose this strategy due to financial or cognitive constraints which persist over time.

not affect the probability ratio between the two first choices. This may not be realistic, especially if μ_2 is close to μ_1. Instead, one would expect the people that preferred the adaptive rule to divide themselves over the two adaptive rules, whereas the people that preferred the fundamentalists rule should be largely unaffected by the introduction of the additional adaptive rule.

Based on the previous assumptions, the probability of choice A at t under the logit model, for the example of two different alternatives, is given by[4]

$$Pr^l_{A,t} = \frac{e^{\frac{\beta}{s}x_{A,t}}}{e^{\frac{\beta}{s}x_{A,t}} + e^{\frac{\beta}{s}x_{B,t}}}, \tag{5}$$

where s is the scaling parameter of the logistic distribution that is related to the variance of ϵ_{it}. This paramter captures the extent of heterogeneity of the unobserved factors for given β. The fraction $\frac{\beta}{s}$, which – in the behavioral finance and macro literature – is known as the *intensity of choice* parameter of the logit-based heuristic switching model. The intensity of choice parameter captures how easily individuals switch between alternative prediction rules.

2.2 Probit

The probit model is able to overcome all of the above mentioned limitations. By assuming that ϵ_t^{ji} are normally distributed, there is no longer a need to assume that these shocks are IID. Hence, different correlation structures between the error terms of different choice alternatives are allowed, and the model is no longer restricted to comply with the IIA property.[5] Moreover, persistent differences between the subjective utility of different people with respect to different alternatives are allowed, and random taste variation can be incorporated.

For the case of two choice alternatives without taste variation (i.e. for the case where $\beta_i = \beta$ is the same across individuals), the probability of an individual choosing option

[4]For a formal proof see Train (2009).
[5]For the theoretical details on this see Train (2009) and Paetz and Steiner (2018).

A under the probit model at t will be given by

$$Pr^{p}_{A,t} = \Phi\left(\frac{1}{\sigma\sqrt{2}}\beta(x_{A,t} - x_{B,t})\right), \tag{6}$$

where σ is the standard deviation of the normal distribution of ϵ^{ji}_t and $\Phi()$ is the cumulative distribution function (CDF) of the standard normal distribution.

A reason that the probit model may not be chosen by a modeler is that the normal CDF needs to be calculated numerically and this puts limits on the analytical results that can be obtained when including probit in a model. However, as we show in Section 5, a number of analytical results can still be obtained on top of the numerical analysis that is more often than not needed for both approaches. Further, depending on the application, the advantages that arise for the probit-model due to the the relaxation of the IID assumption may outweigh and concerns for analytical tractability.

3 Comparing Discrete Choice Models

Before turning to the implications of allowing for random taste variation in Section 4, we first analyze the differences between the logit-based heuristic switching model and a probit-based heuristic switching model without random taste variation. In Section 3.1, we study how well the logit-based and probit-based models can approximate each other when errors are assumed to be IID. Here we put particular focus on the role of the intensity of choice parameter. Next, in Section 3.2, we highlight the role of the independence of irrelevant alternatives property as a driver of differences between the logit- and probit-based heuristic switching models when there are more than two choice alternatives. Finally, in Section 3.3, we show that, depending on the number of choice alternatives, the probit-based model can lead to a considerably better fit to the experimental data of Anufriev et al. (2016) than the standard logit-based heuristic switching model.

3.1 Aligning Behavioral Heterogeneity

The variance of the unobserved characteristics ϵ_t^{ji} captures a type of behavioral heterogeneity across individuals. A high variance, relative to β, means that, in general, more people exist with relatively strong biases towards some choice or another. The ratios $\frac{\beta}{s}$ and $\frac{\beta}{\sigma\sqrt{2}}$, thus, determine the mapping from differences in fitness measures to choice probabilities of the logit-based and probit-based model, respectively. Since the logit-based and probit-based models follow from different distributions, they require different values of the extent of behavioral heterogeneity to imply a similar mapping. Therefore, we will normalize σ and s in such a way that the logit-based and probit-based models imply similar mappings for any common value of β. The latter can then be interpreted as the intensity of choice parameter of both normalized models.

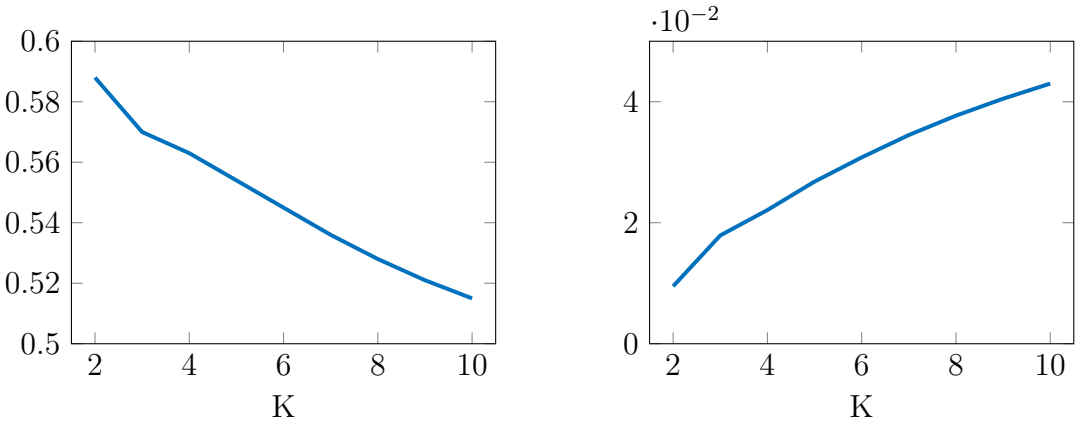
First normalizing $\sigma = \frac{1}{\sqrt{2}}$, we numerically search for the value of s in the logit model that minimizes the maximum absolute difference between the logit- and the probit-based heuristic switching models over the whole support.[6] We do so for the case of $K = 2, \dots 10$ choice alternatives. Further, in order to let the two models be most comparable, we assume for now that the ϵ_t^{ji}'s of the probit model are IID. When calculating the maximum absolute distance, we consider all possible combinations of the fitness measures of all K alternatives.

Panel (a) of Figure 1 shows the resulting values of the distance minimizing scaling factors related to behavioral heterogeneity. For the case of two choice alternatives, we find that the logit- and probit-based models best approximate each other when $\frac{\beta}{s} = 1.702\frac{\beta}{\sigma\sqrt{2}}$, which implies for our normalization of σ that

$$s = \frac{1}{1.702} \approx 0.588. \tag{7}$$

For larger numbers of choice alternatives, s in the multinomial logit model should be

[6]Since the probit and logit mappings for the case of two choice alternatives equal the normal and logistic CDF, our numerical exercise heavily borrows from Bowling et al. (2009) who search for the paramterization for which the logistic CDF best approximates the univariate-normal CDF. In our extended numerical optimization for the case of more than two choice alternatives, we minimize the maximum absolute difference between the two mappings using local optimization routines in combination with a grid of different initial conditions to make sure that the global maximum is found.



(a) Optimal s (b) Maximum distance

Figure 1: Optimal intensity of choice scaling and maximum distance between probit and logit for this intensity of choice scaling for different values of the number of choice alternatives K.

smaller to best approximate the independent multinomial probit model.

In panel (b) of Figure 1, we present the maximum absolute difference that arises between the choice probabilities of the logit-based and probit-based models under the corresponding value of s of panel (a). It can be observed that the largest absolute difference between the models rises when the number of choice alternatives increases. However, for numbers of choice alternatives that are most commonly used in the behavioral finance and macro literature (up to $K = 4$), the largest absolute difference remains relatively small (less than 0.025).

Moreover, for the case of two choice alternatives, the maximum absolute distance even is as low as 0.0095. In Figure 2 we plot the fraction of agents choosing option A as a function of the difference in fitness measures when s satisfies (7). As can be seen in that figure, for any value of the difference in fitness measures between the two choice alternatives, the resulting probability of choosing one of the two options is almost the same under the logit-based as under the probit-based model.

One might, therefore, conclude that, for the case of two choice alternatives and in the absence of random taste variation, the restrictions of the logit-based model do not pose a significant threat, and that the logit model can be seen as a good approximation of the probit-based model. However, there is also a considerable class of behavioral finance models for which a small change in the law of motion of the heuristic switching equation

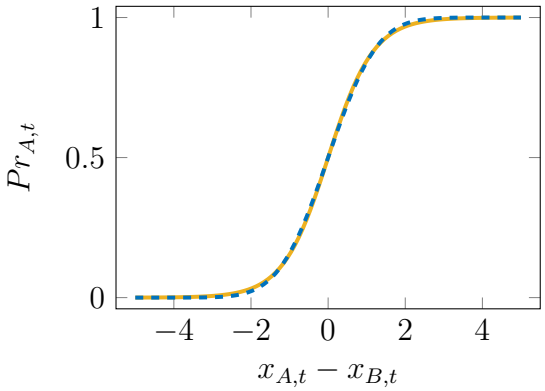


Figure 2: Mapping from difference in fitness measures to probability of choosing an alternative for probit (blue dashed) and logit (orange)

may lead to considerably altered dynamics, not only quantitatively, but even qualitatively. This is especially the case for dynamical systems that may result in chaotic dynamics with sensitive dependence on initial conditions. In Section 5, we focus on the model from the seminal paper of Brock and Hommes (1997) that has these properties and on which many behavioral finance models are based. We show that the main results of that model continue to hold when one assumes the probit model for the heuristic switching mechanism instead of the logit one, but that in a considerable, and relevant, region of the parameter space dynamics are completely altered by the probit assumption.

3.2 The IIA property of logit

Next, we turn to the above mentioned IIA property of logit. Studying this property in some detail will provide more intuition into the differences between logit- and probit-based models for more than two choice alternatives and highlights the flexibility of the latter model.

For the logit-based model, the probability of choosing alternative A at t when there are a total of K alternatives is given by

$$Pr^{l}_{A,t} = \frac{e^{\frac{\beta}{s}x_{A,t}}}{\sum_{j\in J} e^{\frac{\beta}{s}x_{j,t}}}, \tag{8}$$

where J denotes the set of K different choice alternatives.

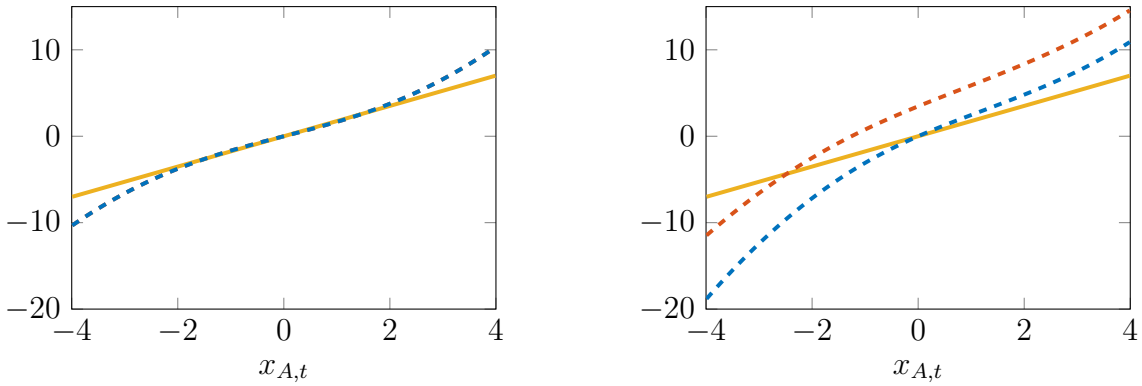
When calculating the ratio between the probabilities of choosing two different choice alternatives (A and B) at time t one then gets

$$\frac{Pr^l_{A,t}}{Pr^l_{B,t}} = \frac{e^{\frac{\beta}{s}x_{A,t}}}{e^{\frac{\beta}{s}x_{B,t}}} = e^{\frac{\beta}{s}\left(x_{A,t}-x_{B,t}\right)}. \tag{9}$$

That is, this ratio does not depend on the fitness measures of any of the other alternatives, no matter how many other alternative choices there are. This is the *independence of irrelevant alternatives* (IIA) property. When the fitness measures of other choice alternatives change, the ratio between the probabilities of choosing alternative A and B will stay the same in the logit model. Moreover, as can be seen in Equation 9, the logarithm of this probability ratio is linear in the fitness measures of the two alternatives.

For the probit model, neither of these two properties hold, even when there is no correlation between the error terms of the choice alternatives. We illustrate this in Figure 3 where we plot the log probability ratio between two choice alternatives for the case of $K = 3$. In particular, we normalize fitness measure of choice alternative B to $x_{B,t} = 0$ and vary the fitness of choice alternative A from -4 to 4. The x-axis can therefore be interpreted as $x_{A,t} - x_{B,t}$. In line with Equation (9), the log probability ratio for the logit model is linear in this difference of fitness measures, as indicated by the upward sloping solid-orange lines. The dashed curves that correspond to the probit model, on the other hand, depict a non-linear relationship between $x_{A,t} - x_{B,t}$ and the log probability ratio. This difference in shapes of the logit and probit log probability ratios is driven by the fact that the normal distribution has smaller tails than the logistic one. Dashed-blue curves in Figure 3 correspond to the independent probit model from the previous subsection, where there is no correlation between the error terms. The red curves (in panel (a) overlapping with blue), depict a case where the error terms of the different alternatives are correlated.[7] The slope of the orange-solid line compared to the dashed curves depends on the value of s which we have set to $s = 0.57$, in line with the optimal value for three choice alternatives from panel (a) of Figure 1.

[7] In particular we take the estimated correlation structure of the SI3 data from Anufriev et al. (2016) that is introduced in the next subsection. The variance covariance matrices that are implied by our estimation are presented in the online appendix.



(a) $x_{C,t} = -10$ (b) $x_{C,t} = 2$

Figure 3: Log probability ratios for $K = 3$ for two different values of $x_{C,t}$. $x_{A,t}$ is varied along the x-axis, and $x_{B,t} = 0$. The solid-orange lines correspond to the logit-based model, whereas dashed curves correspond to probit-based models. In the red case, the error terms of different choice alternatives are correlated, whereas blue corresponds to independent multinomial probit.

Panels (a) and (b) of Figure 3 correspond to two different values of the fitness measure of the third choice alternative. In panel (a) this is set to $x_{C,t} = -10$, making the probability of choosing that alternative effectively zero in both the probit and the logit model. This case is, therefore, practically equivalent to a model with $K = 2$. The difference in the log probability ratios between the logit and the probit model remain here relatively small, and the correlation structure between the error terms has no effect on the log probability ratio.

In panel (b), the fitness measure of the third alternative is, instead, set to $x_{C,t} = 2$ and hence considerably larger than the fitness of alternative B. The IIA property says that the probability ratio (and hence also its log) does not change when the fitness of another alternative is changed. Accordingly, the orange line is in exactly the same position in panel (b) as it was in panel (a). However, the probit model does not satisfy this property and, accordingly, the dashed blue and red curves change when the fitness of the third alternative is changed. As a consequence, the difference between the logit and probit-based models can become larger when the fitness of the third alternative is varied. Furthermore, which combination of fitness measures leads to a larger difference between logit and probit depends on the correlation structure of the error terms in the probit model. By assuming a different correlation structure, choice probabilities will be

affected in a different way by the fitness measures of other alternatives. Note that the latter flexibility property is what makes the crucial difference for applications of the two models. Whereas it is not typically ex-ante clear whether the mappings of logit or of independent probit will provide a better fit to the data, the flexibility of having different mappings for different correlation structure under probit is what really meaningfully relaxes the restrictions of the IIA property.

Finally, we observe that the differences between the blue and orange curves in panel (b) can explain why the maximum absolute difference between the probit- and logit-based model is larger for the case of $K = 3$ than for $K = 2$. Similarly, the different behavior of the two models when other alternatives are varied can also explain why the maximum absolute difference becomes larger and larger as the number of choice alternatives is further increased (as found in panel (b) of Figure 1).[8]

3.3 Fitting Experimental Choice Data: Logit VS Probit

One of the motivations of this paper, has been the laboratory experiment of Anufriev et al. (2016) where experimental human subjects choose between two, three or four profitable alternatives which represent mutual funds, with the time series of the returns of each of the funds having been generated *ex ante*.[9]

Anufriev et al. (2016) find that, while for cases with a small number of choice alternatives, the logit structure is able to capture the behavior shown in the experimental data, the standard logit-based heuristic switching model does not capture the actual choices of the subjects very well when there are four choice alternatives. In order to see whether a probit framework overcomes this limitation, we use the same data and compare the performance of both models.

The probit-based model has the potential to fit actual choice data considerably better, by allowing for correlations between (the error terms of) different choice alternatives. As

[8]Moreover, the steeper slope in the middle part of the dashed curves in panel (b) of Figure 3 compared to panel (a) also hint at an intuition for why a larger K implies a lower optimal value of s. This is because the slope of the orange line is decreasing in s.

[9]The time series in Anufriev et al. (2016) correspond to three types of data: (i) white noise, (ii) data from simulating Brock and Hommes (1997, 1998) and (iii) actual data from stock indices.

discussed above, the logit-based model, instead, features the IID property and, therefore, may not be able to take account of some important aspects of the actual decision problems in practical applications. We note, however, that the probit-based heuristic switching model features more free parameters than the standard logit-based one. Therefore, we compare the performance of both models by considering the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), rather than looking only at the log likelihoods directly. Following Anufriev et al. (2016), we perform maximum likelihood estimations based on a likelihood function that is an aggregation of the individual likelihoods of all subjects in a treatment. These individual likelihood functions represent the likelihoods that a subject would choose the sequence of choice alternative that they ended up choosing in the experiment.

In Tables 1 and 2, we compare our probit-based estimation results with the logit-based findings of Anufriev et al. (2016). Table 1 corresponds to the two data-sets with three choice alternatives and Table 2 depicts the results for the two data-sets where subjects had four choice alternatives. In this section, we focus on the first two columns of each data-set, while the final columns (that correspond to a model with random taste variation) will be discussed in the next section.

For the case of three choice alternatives, the probit-based heuristic switching model has three parameters, two of which determine the correlation between the differences in the errors of the three choice alternatives as well as the relative variances of these error differences. Similarly, for the case of four choice alternatives, there are five parameters that pin down these relations. Similar to the normalizaion of $\sigma = \frac{1}{\sqrt{2}}$ above, in the estimated probit model, we have normalized the first element of the variance covariance matrix of differences in error terms to 1. Details of how these estimated parameters (ϕ_{21}, ϕ_{22}, ...) shape the different variance-covariance matrices are provided in the online appendix. In order to obtain the same parameter estimates for the logit case as Anufriev et al. (2016), we set $s = 1$ in our estimated logit-based heuristic switching model.

Comparing the performance of probit versus logit in Table 1, it can be seen that, for the treatments where subjects made decisions based on actual stock market date (SI3),

Table 1: Maximum likelihood estimation for experimental data with three choice alternatives

	SI3			BH3		
	logit	probit	probit-RT	logit	probit	probit-RT
β	2.008	0.925	2.829	0.039	0.026	0.026
	(0.158)	(0.250)	(1.206)	(0.012)	(0.007)	(0.007)
ϕ_{21}		0.867	1.871		0.582	0.582
		(0.183)	(0.749)		(0.114)	(0.113)
ϕ_{22}		0.655	0.950		0.707	0.707
		(0.290)	(0.348)		(0.123)	(0.122)
σ_β			1.753			0.000
			(0.874)			
log-likelihood	-155.882	-149.664	-140.983	-654.266	-653.501	-653.501
AIC	313.764	305.328	**289.966**	**1310.531**	1313.001	1315.001
BIC	317.399	316.232	**304.505**	**1314.928**	1326.192	1332.589

As in Anufriev et al. (2016), estimations are based on the final 20 experimental periods. The estimations of the columns labeled 'logit' and 'probit' are discussed in Section 3.3. The estimations of the columns labeled 'probit-RT' (random taste variation) are discussed in Section 4. ϕ_{21} and ϕ_{22} pin down the elements of the estimated variance-covariance matrix as discussed in the online appendix.

the log likelihood under the probit-based estimation is considerably less negative than in the logit-based case. On the other hand, for the treatments with the three choice alternatives based on Brock and Hommes (1998) (BH3), the log likelihood under probit is only marginally better than that of probit. As a consequence, the AIC and BIC favor the logit model for the latter case. For the SI3 case, the probit-based model is (slightly) favored by both AIC and BIC.

For the cases with four choice alternatives in Table 2, a quite different picture arises. Here the differences in log likelihood between the probit- and logit-based models are much larger for both SI4 and BH4. Furthermore, even though the probit-based model has five additional estimated parameters, this specification is favored over the logit-based model by both AIC and BIC. For both SI4 and BH4 the margin by which these criteria favor the probit-based model are considerable.

We hence conclude that the ability of the probit model to allow for different correlation structures between the errors of the choice alternatives allows it to potentially fit actual choice data considerably better. For the case of three choice alternatives, we find mixed evidence regarding the added value of the probit-based model. However, for the case of

Table 2: Maximum likelihood estimation for experimental data with four choice alternatives

	SI4			BH4		
	logit	probit	probit-RT	logit	probit	probit-RT
β	2.694	1.075	1.156	0.031	0.041	0.039
	(0.162)	(0.135)	(0.351)	(0.022)	(0.010)	(0.011)
ϕ_{21}		1.079	0.942		0.999	0.998
		(0.137)	(0.245)		(0.018)	(0.001)
ϕ_{22}		0.585	0.321		0.109	0.000
		(0.265)	(0.340)		(0.032)	(0.002)
ϕ_{31}		1.015	0.902		0.999	0.988
		(0.125)	(0.133)		(0.011)	(0.010)
ϕ_{32}		0.499	0.131		0.052	0.000
		(0.126)	(0.677)		(0.016)	(0.003)
ϕ_{33}		0.274	0.184		0.002	0.000
		(0.044)	(0.406)		(0.005)	(0.003)
σ_β			0.198			0.0013
			(0.096)			(0.0007)
log-likelihood	-418.554	-351.672	-348.075	-387.197	-281.588	-268.979
AIC	839.109	715.345	**710.151**	776.395	575.177	**551.957**
BIC	843.437	741.312	**740.447**	780.029	596.985	**577.401**

As in Anufriev et al. (2016), estimations are based on the final 20 experimental periods. The estimations of the columns labeled 'logit' and 'probit' are discussed in Section 3.3. The estimations of the columns labeled 'probit-RT' (random taste variation) are discussed in Section 4. ϕ_{21}, ϕ_{22}, ϕ_{31}, ϕ_{32} and ϕ_{33} pin down the elements of the estimated variance-covariance matrix as discussed in the online appendix.

four choice alternatives we find that the the probit-based heuristic switching model is strongly the preferred model when it comes to matching choices made in a laboratory experiment.[10] Together, these results confirm that the limitations of the logit-based heuristic switching model, as compared to the probit-based one, seem to be of more concern the large the number of potential choice alternatives. As discussed above, this can be explained by the IIA property of the logit model, which, in many applications, is not a realistic assumption.

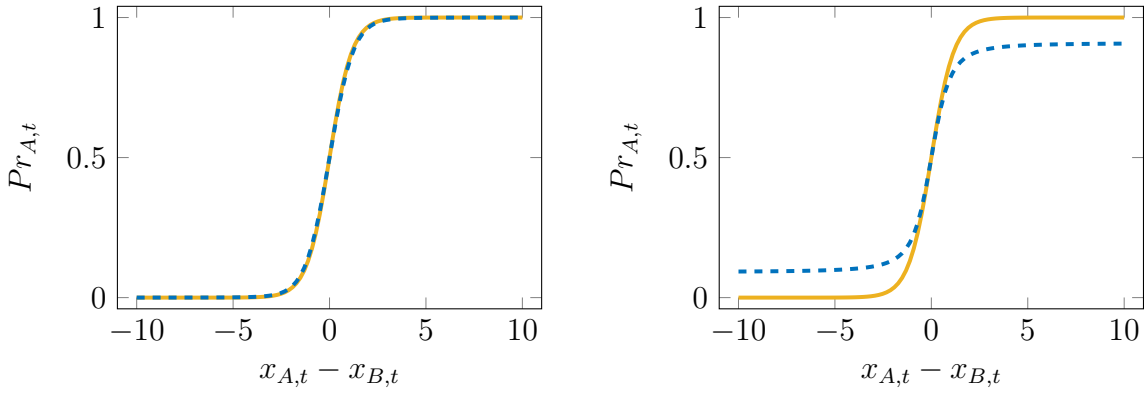
4 Random Taste variation

So far, we have compared logit- and probit-based models under the assumption that the β_i's in Equation (2) are the same for all individuals. As mentioned in Section 2, one of the advantages of the probit model is that it allows for random taste variation across agents. In particular, the β_i's are allowed to be normally distributed across individuals with some mean $\bar{\beta}$ and some variance-covariance matrix. The utility function of agent i, can then be rewritten as in (4), i.e. $U_t^{ji} = \bar{\beta} x_{j,t} + \eta_t^{ji}$, with $\eta_t^{ji} = \tilde{\beta}_i x_{j,t} + \epsilon_t^{ji}$ and $\tilde{\beta}_i = \beta_i - \bar{\beta}$. Then, under the assumption that the β_i's are normally distributed, η_t^{ji} will still be normally distributed, but with a different variance-covariance matrix than that of ϵ_t^{ji}. Since the probit framework does not require η_t^{ji} to be IID, this model specification can be used for estimating or simulating a probit model.

Consider the case of a single observable variable to measure performance, so that β_i and $x_{j,t}$ are scalars. Letting σ_β^2 be the variance of the normally distributed β_i's, we then have $\beta_i \sim \mathcal{N}(\bar{\beta},\, \sigma_\beta^2)$ and $\tilde{\beta}_i \sim \mathcal{N}(0,\, \sigma_\beta^2)$. Further using that $\epsilon_t^{ji} \sim \mathcal{N}(0,\, \sigma^2)$, the variance-covariance matrix of η_t^{ji} for the case of two choice alternatives becomes

$$\Omega = \sigma_\beta^2 \begin{bmatrix} x_{A,t}^2 & x_{A,t}x_{B,t} \\ x_{A,t}x_{B,t} & x_{B,t}^2 \end{bmatrix} + \sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{10}$$

[10] We also estimated our probit-based heuristic switching model for the case of two choice alternatives. Here, we not only find mixed results in terms of AIC and BIC but also with regard to the log likelihood of the logit-based and probit- based models directly. This is because the probit-based model does not feature any additional parameters in the case of two choice alternatives. Full results are presented in Table 3 in Appendix B.



(a) $\sigma_\beta = 0.25$ (b) $\sigma_\beta = 0.75$

Figure 4: Mapping from difference in fitness measures to probability of choosing an alternative for logit (solid) and probit with random taste variation (dashed) for two different levels of heterogeneity in tastes.

Hence, the difference between η_t^{ji}'s is distributed as $\eta_t^{Ai} - \eta_t^{Bi} \sim \mathcal{N}(0,\ (x_{A,t} - x_{B,t})^2\sigma_\beta^2 + 2\sigma^2)$. Therefore, the random-taste equivalent of (6) can be written as

$$Pr_{A,t}^{p} = \Phi\left(\frac{\bar{\beta}}{\sqrt{(x_{A,t} - x_{B,t})^2\sigma_\beta^2 + 2\sigma^2}}(x_{A,t} - x_{B,t})\right). \tag{11}$$

This means that, even if we normalize for scale by setting $\sigma = \frac{1}{\sqrt{2}}$, the switching intensity still is affected by two parameters: σ_β and $\bar{\beta}$.

Figure 4 presents the mapping from differences in fitness measures to the probability of choosing heuristic A under random taste variation for the case that $\bar{\beta} = 1$. The solid orange lines present the logit case without random taste variation and with $s = 0.588$, just as in Figure 2. Panel (a) depicts the case of $\sigma_\beta = 0.25$, whereas panel (b) corresponds to $\sigma_\beta = 0.75$.

In Panel (a) of Figure 4, the dashed-blue curve of the probit-based model always remains very close to the solid logit curve. Apart from the different range of values displayed on the x-axis, this panel is practically indistinguishable from Figure 2. However, when the heterogeneity in tastes is increased further, as in panel (b), an important qualitative differences arises in the probit based-model with random taste variation. In particular, the dashed curve in panel (b) no longer converges to 0 and 1 when the difference in fitness measures goes to minus infinity or infinity respectively. This leads to a considerable dif-

ference between the mappings of fitness measure to choice probabilities of the logit-based and probit-based models when differences in fitness measures are not close to zero.

The intuition for this result is that, for $\sigma_\beta = 0.75$ and $\beta = 1$, some agents will draw a β_i around 0, or even a negative β_i. These agents become completely indifferent to fitness measures or even prefer alternatives with lower $x_{j,t}$ over alternatives with a higher $x_{j,t}$. Therefore, a small fraction of agents will still choose a heuristic that has a considerably smaller fitness measure than the alternative. Therefore, no matter how large the difference in fitness measures, fractions of 0 or 1 are never reached.

Next, we consider whether allowing for random taste variation can further improve the fit of the probit-based heuristic switching model to the experimental data of Anufriev et al. (2016), compared to the estimates discussed in the previous section. The estimation results of random-taste probit ('probit-RT') are presented next to their logit and regular-probit counterparts in Tables 1 and 2.

For the case of three choice alternatives in Table 1, we again obtain mixed results. For the SI3 model there is a relevant role for random taste variation, and, with just one extra parameter, the log likelihood is improved considerably compared to the probit model without random taste variation. In terms of AIC and BIC, random-taste probit is hence clearly the preferred model, also compared to the logit-based heuristic switching model. For BH3, on the on the other hand, there seems to be no role for random taste variation in the experimental data. Here σ_β is estimated to be 0, and the likelihood is exactly the same as in the probit model without random taste variation.

At the same time, in Table 2, we find a further improvement of the log-likelihood as well as the AIC and BIC for both data-sets with four choice alternatives when random taste variation is allowed for. All in all, the above underlines further advantages of the probit-based framework and shows that random taste variation is an empirically relevant phenomenon. However, this seems to depend on the particular choice problem considered.

5 A Probit-based Cobweb Model

In this final section, we turn to a theoretical application of the probit-based heuristic switching model where we abstract from random-taste variation. Moreover, in this application, there only are two choice alternatives, so that correlations of errors of different choice aternatives do not play a role. The purpose is to show that although most results are robust to assuming probit instead of logit, important qualitative differences may arise, even in this case where the logit-based and probit-based models are most similar (see Figure 2).

In particular, we investigate the adaptive rational equilibrium dynamics (A.R.E.D.) that arise in the probit-based version of the cobweb model of Brock and Hommes (1997) (henceforth BH) and we compare our results with those of the logit-based model of BH. Assume an economy populated by a large number of heterogeneous agents who, in each period, produce an asset. The price P_t of the asset at time t is revealed at the end of the period based on agents' total supply and on the demand of the asset.

As in BH, agents can choose between two predictors, $H_j, j \in \{P, N\}$, to make predictions about the price. Here, predictor H_P corresponds to perfect foresight and predictor H_N corresponds to naive, or static, expectations:

$$H_P(\boldsymbol{P}_t) = P_{t+1}, \tag{12}$$

$$H_N(\boldsymbol{P}_t) = P_t, \tag{13}$$

where $\boldsymbol{P}_t$ is a vector of known past prices, $\boldsymbol{P}_t = (P_t, P_{t-1}, P_{t-2}, ..., P_{t-L})$, extending $L \in \mathbb{N}$ periods back. In order to utilize predictor H_P during a period t, agents need to pay a fixed positive information cost, $C \in [0, \infty)$. There is no such cost incurred by agents who utilize the naive expectations predictor, H_N.

Demand and supply are linear and are given by

$$D(P_t) = A - BP_t \tag{14}$$

with $A, B \in (0, \infty)$ and

$$S\left[H_j(\boldsymbol{P}_t)\right] = bH_j(\boldsymbol{P}_t), \tag{15}$$

where the supply function is given by agents' maximization of profits, assuming quadratic costs per output q equal to $q^2/2b$, with $b \in (0, \infty)$.

The performance measures for predictors is net realized profits. These are derived in BH to be

$$x_{P,t} = \pi_P(P_{t+1}) = \frac{b}{2}P_{t+1}^2 - C, \tag{16}$$

for the perfect foresight predictor and

$$x_{N,t} = \pi_N(P_{t+1}, P_t) = \frac{b}{2}P_t(2P_{t+1} - P_t), \tag{17}$$

for the naive expectations predictor.

The evolution of the price, given by the equalization of supply and demand is the same as in BH and is given by

$$P_{t+1} = -\frac{b(1 - m_t)P_t}{2B + b(1 + m_t)}. \tag{18}$$

Plugging (16) and (17) into (6) gives us the fractions $n_{P,t+1}$ and $n_{N,t+1}$ of agents following each of the two predictors under the probit assumption. After some rearrangement and normalizing $\sigma = \frac{1}{\sqrt{2}}$, the difference in the fractions of agents following the two predictors can be written as[11]

$$m_{t+1} = n_{P,t+1} - n_{N,t+1} = 2\Phi\Big(\beta\Big[\frac{b}{2}\Big(\frac{b(1 - m_t)P_t}{2B + b(1 + m_t)} + P_t\Big)^2 - C\Big]\Big) - 1. \tag{19}$$

Hence, the final form of the second-order dynamics for P_t and m_t is given by (18) and (19).

The following Proposition corresponds to the case of no costs for having the perfect foresight predictor ($C = 0$). The proofs of all Propositions are provided in Appendix A.

[11] The derivation is given in the online appendix.

Proposition 1. *Consider the economy described by* (18) *and* (19). *Then, if* $C = 0$, $(\bar{P}, \bar{m}) = (0, 0)$ *is the unique fixed point and it is always globally asymptotically stable.*

Next, we will turn to the more interesting case of positive costs, C. Before we present the characterization of the fixed points for that case, we define two different types of fixed points.

Definition 1. *Let* $\underline{f}(x, y) = (f_1(x,\ y),\ f_2(x,\ y))$ *be a map on* $\mathbb{R}^2$.
Assume that $\Big(f_1(p_1, p_2), f_2(p_1, p_2)\Big) = (p_1, p_2)$ *is a fixed point.*

(i) If all of the eigenvalues of the Jacobian matrix $\left.\dfrac{\partial(f_1, f_2)}{\partial(x, y)}\right|_{(p_1,p_2)}$ *are, in absolute value, smaller than 1, then the fixed point is asymptotically stable.*

(ii) If one of the eigenvalues of the Jacobian matrix $\left.\dfrac{\partial(f_1, f_2)}{\partial(x, y)}\right|_{(p_1,p_2)}$ *is, in absolute value, greater than 1 and the other eigenvalue smaller than 1, then the fixed point is called a saddle and is unstable.*

Now, let $\beta^\star = \dfrac{\sqrt{2} erf^{-1}\left(\dfrac{B}{b}\right)}{C}$.

Proposition 2. *Let* $C \in (0, \infty)$, *then:*

(i) For each $\beta \geq 0$ *the fixed point is* $(\bar{P}, \bar{m}) = \left(0, -erf\left(\dfrac{\beta C}{\sqrt{2}}\right)\right)$.

(ii) At $\beta = \beta^\star$ *a period-doubling bifurcation occurs such that the fixed point is globally stable for* $0 \leq \beta < \beta^\star$ *and the fixed point is a saddle for* $\beta > \beta^\star$.

The first part of Proposition 2 shows that when the costs of choosing the perfect foresight predictor are positive, then the fixed point value of the relative share of agents choosing the perfect foresight – compared to the naive one – depends negatively both on the costs and on the intensity of choice parameter. Given that, at the fixed point, the two predictors give the same predictions, it is intuitive that the relative share depends on the costs of choosing the perfect foresight one. The exact fraction that still chooses the perfect foresight predictor for a given value of its costs then depends on the intensity of choice parameter, in line with the discussions of this parameter in Sections 2 and 3.

The next part of the proposition focuses on the relation between the intensity of choice parameter and the asymptotic stability properties of the fixed point. When the intensity of choice parameter is low, the fixed point is locally stable, while when it crosses a threshold value, the fixed point becomes locally unstable (a saddle). The interesting point here is that, as β captures agents' level of rationality, instability emerges due to agents being sufficiently rational. Obviously, this is due to the costs of using the foresight predictor. Higher β means that more agents would choose the perfect foresight predictor when profitable. However, this implies less profits, and, given that costs of having the perfect foresight predictor are fixed, the predictor becomes less profitable. Hence, the force that leads to agents choosing one predictor due to the related profits is the same as the one that makes the predictor less profitable in future periods and this is the source of instability. In the limiting case of $\beta \to \infty$, $(\bar{P}, \bar{m}) = (0, -1)$ which means that the optimal predictor at the fixed point is the naive one for all agents. However, as noted above, this would not be a stable regime.

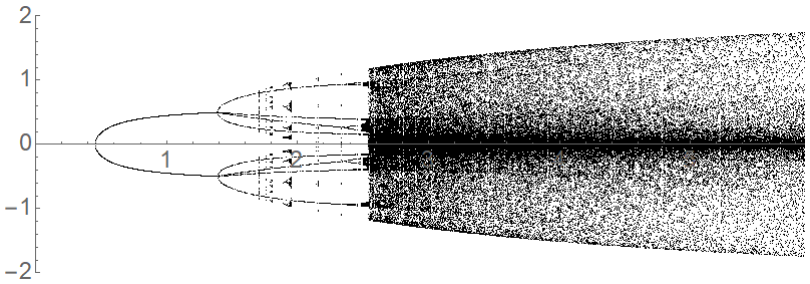
Further, observe that the period-doubling bifurcation presented in Proposition 2 implies the creation of a 2-cycle. It turns out that this 2-cycle is initially locally stable, but then loses stability in another bifurcation. This is formally presented in Proposition 3.

Proposition 3. *Let $C \in (0, \infty)$, then there exists a β^{**}:*

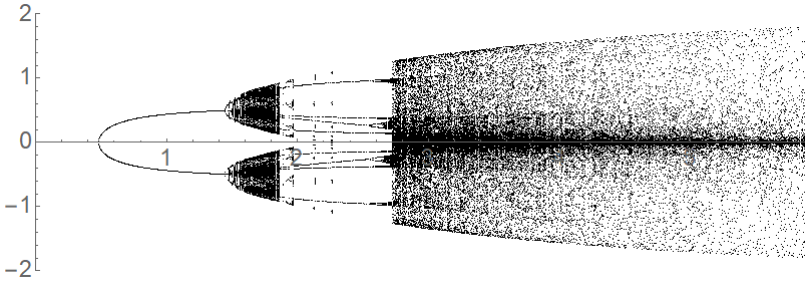
(i) For $\beta^{\star} < \beta < \beta^{\star\star}$, there exists a locally stable 2-cycle.

(ii) At $\beta = \beta^{\star\star}$, a Hopf bifurcation occurs such that the 2-cycle is unstable for $\beta > \beta^{\star\star}$.

Proposition 3 describes a second bifurcation where the two-cycle moves to instability as the system's eigenvalues become complex. This is similar to BH Theorem 3.4, with one key difference. While the analytical results presented in Propositions 1, 2 and 3 are qualitatively the same as in BH (Theorem 3.1 and parts (i) and (ii) of 3.4), we have not shown that after the second bifurcation when the 2-cycle loses stability a four-cycle exists (BH, Theorem 3.4 (iii)). Hence, we can only conclude from the above that the behavior of the two approaches is qualitatively the same up to the secondary bifurcation. As we show below with a numerical analysis, there are some key differences after that, with



(a) Logit



(b) Probit

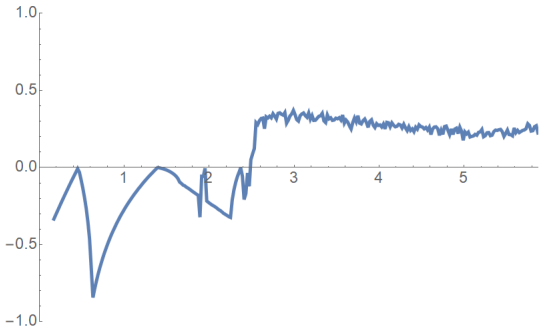
Figure 5: Bifurcation diagrams in β

the probit-based model featuring an additional region with chaotic behavior instead of a four-cycle.

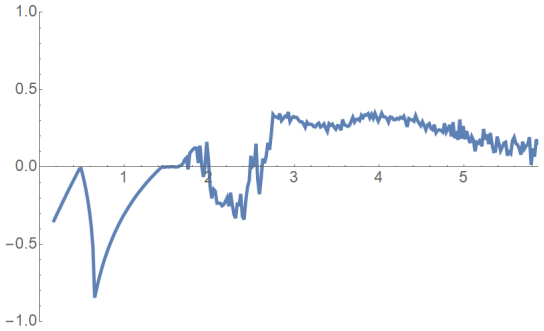
For the numerical analysis, we start by simulating the asymptotic behavior of the model for a range of values of the bifurcation parameter β (bifurcation diagram) and we compare this with the logit-based version of BH. Here, we have normalized s in the logit specification as in (7) in order to make the two models comparable.

We note two interesting features. First, we observe very similar behaviors in the two models, including the same initial region of local stability of the steady state (as indicated by a single solid line at $\bar{P} = 0$) followed by a two-cycle (indicated by two curves symmetric around the $\bar{P} = 0$ fixed point), in line with the above propositions. Moreover, under our normalization, bifurcations in both model specifications occur at similar values of β. For example, for the probit-based model we find $\beta^{\star} \approx 0.48$, whereas for the logit-specification the first bifurcation occurs at $\beta \approx 0.47$.

At the same time, the second interesting feature that can be observed in the bifurcation diagrams is that in the probit framework an extra 'dark' region is created, directly after the region where a 2-cycle occurs. That is, directly after the Hopf bifurcation ($\beta^{\star\star}$)



(a) Logit



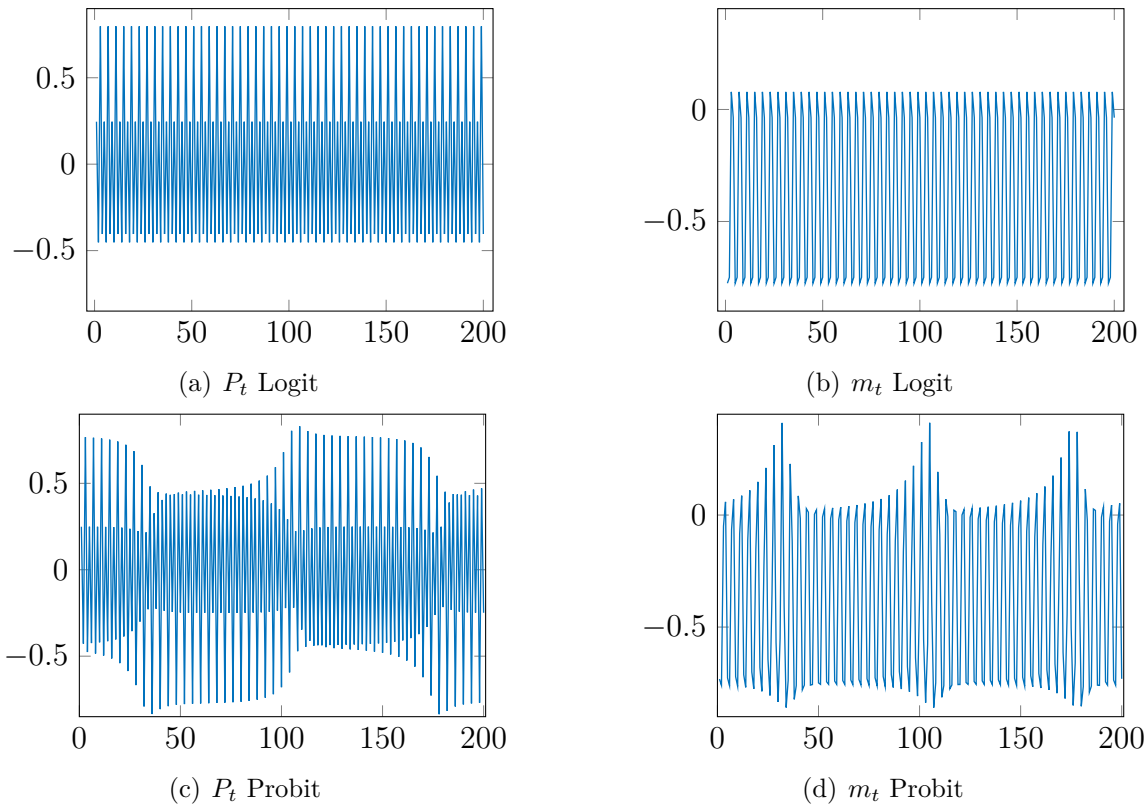
(b) Probit

Figure 6: Largest Lyapunov exponents as function of β

of Proposition 3. To investigate, whether this means that the probit-based models features different results regarding the occurrence of chaotic dynamics, we plot the largest Lyapunov exponents as a function of the intensity of choice for both models in Figure 6.

In both panels, the largest Lyapunov exponent is positive in the right part of the graph, which indicates that there are chaotic dynamics in that region of the parameter space. The well known chaotic behavior for large intensity of choice, which is a key charachteristic of the model in BH, hence, can be said to be robust to assuming a probit framework.

However, for the probit-based model, there is an additional region with positive Lyapunov exponent already for lower intensity of choice. This implies that chaos also already appears in the extra 'dark' region in the bifurcation diagram, directly after the second bifurcation. Since chaos now already arises for much lower value of the intensity of choice, we consider this to be an important qualitative difference between the probit- and the logit-based versions of the model. Note, though, that in the probit model, chaos first disappears again when the intensity of choice is increased to values above the new chaos



(a) P_t Logit (b) m_t Logit

(c) P_t Probit (d) m_t Probit

Figure 7: Time series simulations

region, and then reappears at a similar intensity of choice value as in the logit model.

Next, we investigate the properties and dynamics of the model in this additional chaos region somewhat more closely. Figure 7 shows the the dynamics of P_t an m_t for the probit and logit model for $\beta = 1.7$.[12] For the logit-based model, a clear 4-cycle can be observed (especially when looking at the price dynamics). Moreover, there are periods (e.g. periods 120-160) where the dynamics of the probit-based model are very similar to those of the logit-based one. However, in other periods (e.g periods 50-80) the price dynamics of the probit-based model seem to be inverted compared to the logit case. Furthermore, there are transition periods (e.g periods 80-120) where both P_t and m_t take on a range of values that considerably differ from the regular dynamics in the logit-based model.

Finally, we simulate the model for 1,000,000 periods and plot the simulated points in P, m-space in Figure 8.[13] The top panel features 4 points. This is in line with the 4-cycles that arise in the logit model. For the probit case in the bottom panel, we observe

[12]We throw away the first 20,000 observations of the simulations to prevent dependence on initial conditions and show the subsequent 200 periods which are representative for all subsequent periods.

[13]We again remove the effects of initial conditions by not displaying the firs 20,000 periods

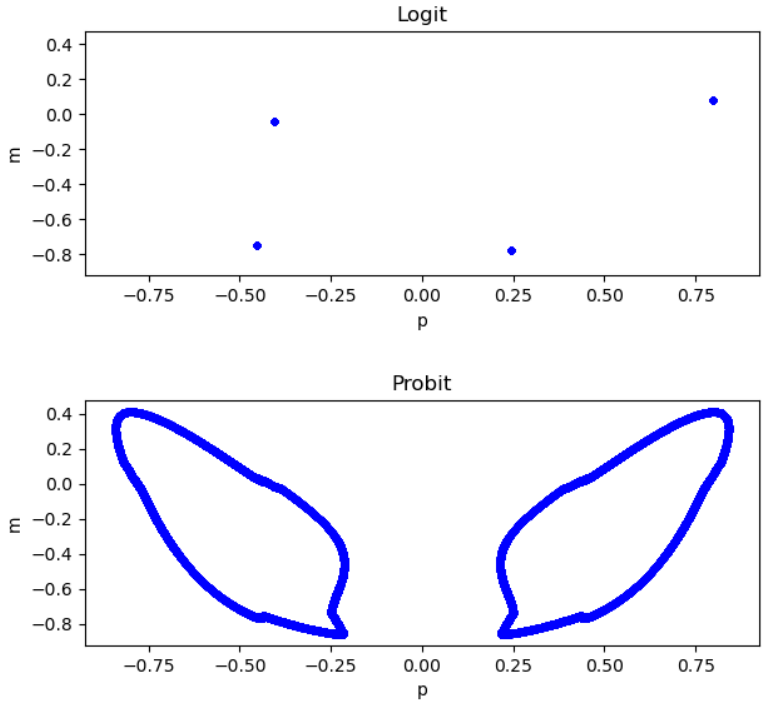


Figure 8: Strange attractor and 4-cycle in P, m-plane

a strange attractor where a continuum of points is reached during the simulations. This figure therefore clearly illustrates that in a region of relatively low intensity of choice, where under logit regular 4-cycles occur, the probit assumptions instead results in a strange attractor with chaotic dynamics.

We conclude that most qualitative properties of the logit-based version of the cobweb model with two alternatives remain when we instead assume a probit-based specification. However, there are also important qualitative differences between the two models in a relevant region of the parameter space, with the probit framework leading to additional chaotic behavior. This highlights that small differences related to the (de)tails of the different distributions – which in economic terms are related to assumptions regarding heterogeneity and preferences over alternatives – can lead to significant changes in the behavior of the model.

6 Conclusion

The behavioral finance and behavioral macroeconomics literature using dynamic discrete choice models is growing in a number of directions. However, the modelling approach related to the random utility framework exclusively assumes that the random element follows an extreme value distribution. As we discuss here, even though this assumption leads to a logit structure of the switching mechanism which is useful in terms of offering analytical tractability, this comes with limitations with regards to the economic environments that it can represent, which is also reflects in limitations as regards to matching experimental data. Even though a probit framework is more general, it has been unclear whether it is able to fit better the data than logit and what the implications of using a probit structure would be in other models. This has been the motivation of this paper.

We first show that, whereas in the case of two alternatives the two approaches are very similar, their differences become more obvious with higher number of choice alternatives. The independence of irrelevant alternatives property of logit plays an important role here. We then find that the probit framework is able to better fit the data with more than two choice alternatives of Anufriev et al. (2016) who found the standard logit model not to be convincing there. Additionaly, we find that, for most time series, the probit model performs even better (in terms of AIC and BIC) if random taste variation is allowed for. Finally, we study the behavior of the BH heterogeneous cobweb asset pricing model under the more general probit assumption and show that while the key results hold, the probit framework leads to a more rich behavior with one more region of chaos appearing.

Our paper raises the point that future works should also consider the robustness of the results under alternative economic assumptions regarding variation in preferences for example. In particular, our work opens the door for many different extensions. In an ongoing follow-up project, we investigate the implications of allowing for random taste variation in the BH model. The same can be done in other models from the literature, including macroeconomic models with heuristic switching, such as in Hommes and Lustenhouwer (2019). At the same time, the implications of different correlation structures between choice alternatives for financial and macroeconomic models that fea-

ture heuristic switching with more than two choice alternatives can be investigated in a probit-framework. Furthermore, the probit based heuristic switching model can be estimated on a wealth of other experimental data from various existing learning-to-forecast experiments.

References

Anufriev, M., Bao, T., Tuinstra, J., 2016. Microfoundations for switching behavior in heterogeneous agent models: An experiment. Journal of Economic Behavior & Organization 129, 74–99.

Anufriev, M., Hommes, C., 2012. Evolutionary selection of individual expectations and aggregate outcomes in asset pricing experiments. American Economic Journal: Microeconomics 4, 35–64.

Anufriev, M., Hommes, C., Makarewicz, T., 2018. Simple Forecasting Heuristics that Make us Smart: Evidence from Different Market Experiments. Journal of the European Economic Association 17, 1538–1584.

Assenza, T., Heemeijer, P., Hommes, C.H., Massaro, D., 2021. Managing self-organization of expectations through monetary policy: A macro experiment. Journal of Monetary Economics 117, 170–186.

Bao, T., Hommes, C., Makarewicz, T., 2017. Bubble Formation and (In)Efficient Markets in Learning-to-Forecast and optimise Experiments. The Economic Journal 127, F581–F609.

Bao, T., Hommes, C., Sonnemans, J., Tuinstra, J., 2012. Individual expectations, limited rationality and aggregate outcomes. Journal of Economic Dynamics and Control 36, 1101–1120.

Bordalo, P., Gennaioli, N., Shleifer, A., 2018. Diagnostic expectations and credit cycles. The Journal of Finance 73, 199–227.

Bowling, S.R., Khasawneh, M.T., Kaewkuekool, S., Cho, B.R., 2009. A logistic approximation to the cumulative normal distribution. Journal of Industrial Engineering and Management 2(1), 114–127.

Branch, W.A., 2004. The theory of rationally heterogeneous expectations: Evidence from survey data on inflation expectations*. The Economic Journal 114, 592–621. URL: http://dx.doi.org/10.1111/j.1468-0297.2004.00233.x, doi:10.1111/j.1468-0297.2004.00233.x.

Branch, W.A., McGough, B., 2018. Chapter 1 - heterogeneous expectations and microfoundations in macroeconomics, in: Hommes, C., LeBaron, B. (Eds.), Handbook of Computational Economics. Elsevier. volume 4 of *Handbook of Computational Economics*, pp. 3–62.

Brock, W., Durlauf, S., 2001. Discrete choice with social interactions. Review of Economic Studies 68, 235–260.

Brock, W., Hommes, C., 1997. A rational route to randomness. Econometrica 65, 1059–1095.

Brock, W., Hommes, C., 1998. Heterogeneous beliefs and routes to chaos in a simple asset pricing model. Journal of Economic Dynamics and Control 22, 1235–1274.

Chiarella, C., Dieci, R., He, X.Z., 2009. Chapter 5 - heterogeneity, market mechanisms, and asset price dynamics, in: Hens, T., Schenk-Hoppé, K.R. (Eds.), Handbook of Financial Markets: Dynamics and Evolution. North-Holland, San Diego. Handbooks in Finance, pp. 277–344.

Chiarella, C., He, X., 2002. Heterogeneous beliefs, risk and learning in a simple asset pricing model. Computational Economics 19, 95–132.

Chiarella, C., He, X., 2003. Heterogeneous beliefs, risk and learning in a simple asset pricing model with a market maker. Macroeconomic Dynamics 7, 503–536.

Colasante, A., Palestrini, A., Russo, A., Gallegati, M., 2017. Adaptive expectations versus rational expectations: Evidence from the lab. International Journal of Forecasting 33, 988–1006.

Cornea-Madeira, A., Hommes, C., Massaro, D., 2019. Behavioral heterogeneity in us inflation dynamics. Journal of Business & Economic Statistics 37, 288–300.

De Bondt, W.F.M., Thaler, R., 1985. Does the stock market overreact? The Journal of Finance 3, 793–805.

De Grauwe, P., 2011. Animal spirits and monetary policy. Economic Theory 47, 423–457.

De Grauwe, P., 2012. Booms and busts in economic activity: A behavioral explanation. Journal of Economic Behavior & Organization 83, 484–501.

De Grauwe, P., Grimaldi, M., 2005. Heterogeneity of agents, transaction costs and the exchange rate. Journal of Economic Dynamics and Control 29, 691–719.

De Long, J.B., Shleifer, A., Summers, L.H., Waldmann, R.J., 1990. Noise trader risk in financial markets. Journal of Political Economy 98, 703–738.

Di Guilmi, C., Galanis, G., 2021. Convergence and divergence in dynamic voting with inequality. Journal of Economic Behavior & Organization 187, 137–158.

Di Guilmi, C., Galanis, G., Baskozos, G., 2020. A Behavioural SIR Model and its Implications for Physical Distancing. Technical Report. CRETA Online Discussion Paper Series 58.

Dieci, R., He, X.Z., 2018. Chapter 5 - heterogeneous agent models in finance, in: Hommes, C., LeBaron, B. (Eds.), Handbook of Computational Economics. Elsevier. volume 4 of *Handbook of Computational Economics*, pp. 257–328.

Evans, G.W., Honkapohja, S., 2012. Learning and expectations in macroeconomics. Princeton University Press.

Flaschel, P., Charpe, M., Galanis, G., Proano, C.R., Veneziani, R., 2018. Macroeconomic and stock market interactions with endogenous aggregate sentiment dynamics. Journal of Economic Dynamics and Control 91, 237 – 256.

Gabaix, X., 2020. A behavioral new keynesian model. American Economic Review 110, 2271–2327.

Heemeijer, P., Hommes, C., Sonnemans, J., Tuinstra, J., 2009. Price stability and volatility in markets with positive and negative expectations feedback: An experimental investigation. Journal of Economic Dynamics and Control 33, 1052–1072. Complexity in Economics and Finance.

Hommes, C., 2006. Heterogeneous agent models in economics and finance, in: Tesfatsion, L., Judd, K. (Eds.), Handbook of Computational Economics, Vol. 2: Agent-Based Computational Economics. North-Holland, Amsterdam, pp. 1109–1186.

Hommes, C., 2021. Behavioral and experimental macroeconomics and policy analysis: A complex systems approach. Journal of Economic Literature 59, 149–219.

Hommes, C., Lustenhouwer, J., 2019. Inflation targeting and liquidity traps under endogenous credibility. Journal of Monetary Economics 107, 48–62.

Hommes, C., Lustenhouwer, J., Mavromatis, K., 2018. Fiscal consolidations and heterogeneous expectations. Journal of Economic Dynamics and Control 87, 173–205.

Hommes, C., Massaro, D., Weber, M., 2019. Monetary policy under behavioral expectations: Theory and experiment. European Economic Review 118, 193 – 212.

Hommes, C., Sonnemans, J., Tuinstra, J., van de Velden, H., 2005. Coordination of Expectations in Asset Pricing Experiments. The Review of Financial Studies 18, 955–980.

Kahneman, D., 2003. Maps of bounded rationality: Psychology for behavioral economics. American Economic Review 93, 1449–1475.

Kahneman, D., Tversky, A., 1973. On the psychology of prediction. Psychological Review 80, 237–251.

Kopányi-Peuker, A., Weber, M., 2020. Experience Does Not Eliminate Bubbles: Experimental Evidence. The Review of Financial Studies 34, 4450–4485.

Lux, T., 2009. Chapter 3 - stochastic behavioral asset-pricing models and the stylized facts, in: Hens, T., Schenk-Hoppé, K.R. (Eds.), Handbook of Financial Markets: Dynamics and Evolution. North-Holland, San Diego. Handbooks in Finance, pp. 161–215. URL: https://www.sciencedirect.com/science/article/pii/B9780123742582500075, doi:https://doi.org/10.1016/B978-012374258-2.50007-5.

Manski, C., McFadden, D. (Eds.), 1981. Structural analysis of discrete data with econometric applications. Cambridge University Press, Cambridge MA.

McFadden, D., 1974. The measurement of urban travel demand. Journal of Public Economics 3, 303–328.

McFadden, D., 2001. Economic Choices. American Economic Review 91, 351–378.

Paetz, F., Steiner, W.J., 2018. Utility independence versus iia property in independent probit models. Journal of Choice Modelling 26, 41–47.

Shleifer, A., Vishny, R.W., 1997. The limits of arbitrage. The Journal of Finance 52, 35–55.

Azeredo da Silveira, R., Woodford, M., 2019. Noisy memory and over-reaction to news, in: AEA Papers and Proceedings, pp. 557–61.

Simon, H.A., 1957. Models of man. Wiley, NY.

Simon, H.A., 1979. Rational decision making in business organizations. American Economic Review 69, 493–513.

Train, K., 2009. Discrete Choice Methods with Simulation. Cambridge University Press, Princeton.

Westerhoff, F.H., Dieci, R., 2006. The effectiveness of keynes–tobin transaction taxes when heterogeneous agents can trade in different markets: A behavioral finance approach. Journal of Economic Dynamics and Control 30, 293–322.

Woodford, M., 2013. Macroeconomic analysis without the rational expectations hypothesis. Annu. Rev. Econ. 5, 303–346.

Appendix

A Proofs of Propositions

Proof of Proposition 1

At the fixed point $(\bar{P}, \bar{m})$, (18) and (19) become

$$\bar{P} = -\frac{b(1-\bar{m})\bar{P}}{2B + b(1+\bar{m})}, \tag{A.1}$$

and

$$\bar{m} = 2\Phi\Big(\beta\Big[\frac{b}{2}\Big(\frac{b(1-\bar{m})\bar{P}}{2B + b(1+\bar{m})} + \bar{P}\Big)^2 - C\Big]\Big) - 1 \tag{A.2}$$

from which it is apparent that $(\bar{P}, \bar{m}) = (0, 0)$ is a fixed point when $C = 0$. We first prove local stability.

We know that for any $\underline{f}(x, y) = (f_1(x,\ y),\ f_2(x,\ y)\ )$ which is a map on $\mathbb{R}^2$ and a fixed point $\Big(f_1(p_1, p_2), f_2(p_1, p_2)\Big) = (p_1, p_2)$, if all eigenvalues of the Jacobian matrix $\dfrac{\partial(f_1, f_2)}{\partial(x, y)}\Big|_{(p_1,p_2)}$ have magnitude less than 1, then (p_1, p_2) is a fixed-point sink.

The eigenvalues of the Jacobian matrix $\dfrac{\partial(f, g)}{\partial(P, m)}\Big|_{(0,0)}$ are 0 and $-\dfrac{b}{2B + b}$ both of which have magnitude less than 1. This implies that the origin is a hyperbolic fixed-point sink. This proves local stability of the origin.

In order to prove global stability set $A_\mathrm{t} = -\dfrac{b(1-m_\mathrm{t})}{2B + b(1+m_\mathrm{t})}$. Then the price dynamics given by (18) can be written as

$$P_{\mathrm{t}+1} = A_\mathrm{t} P_t.$$

The difference between the fractions of agents choosing between predictors H_P and H_N is less than one in absolute value, i.e., $|m_\mathrm{t}| < 1$.

The function $A_\mathrm{t} = -\dfrac{b(1-m_\mathrm{t})}{2B + b(1+m_\mathrm{t})}$ is a hyperbola with negative values for all $|m_\mathrm{t}| < 1$, is strictly increasing, and its curve has a horizontal asymptote given by the line $A_\mathrm{t} = 1$ Therefore, as long as $|A_\mathrm{t}| = \Big| -\dfrac{b(1-m_\mathrm{t})}{2B + b(1+m_\mathrm{t})}\Big| < 1$ this implies directly that $P_\mathrm{t} \to 0$ as $t \to \infty$ which in turn implies that $m_\mathrm{t} \to 0$.

□

Proof of Proposition 2

(i) If $C \in (0, \infty)$, $\bar{P} = 0$, and $\bar{m}$ denotes the steady state value of the state variable m_t, then (19) becomes

$$\bar{m} = 2\Phi(-\beta C) - 1, \tag{A.3}$$

or equivalently

$$\bar{m} = 1 - 2\Phi(\beta C) \tag{A.4}$$

where

$$\Phi(\beta C) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta C} e^{-\frac{z^2}{2}} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{z^2}{2}} dz + \frac{1}{\sqrt{2\pi}} \int_{0}^{\beta C} e^{-\frac{z^2}{2}} dz = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_{0}^{\beta C} e^{-\frac{z^2}{2}} dz,$$

implies

$$\bar{m} = 1 - 2\left(\frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_{0}^{\beta C} e^{-\frac{z^2}{2}} dz\right) = -\frac{\sqrt{2}}{\sqrt{\pi}} \int_{0}^{\beta C} e^{-\frac{z^2}{2}} dz. \tag{A.5}$$

Let $\frac{z}{\sqrt{2}} = u$ such that $dz = \sqrt{2}du$. Then, (A.5) becomes

$$\bar{m} = -\frac{\sqrt{2}}{\sqrt{\pi}} \int_{0}^{\frac{\beta C}{\sqrt{2}}} e^{-u^2} \sqrt{2} du = -erf\left(\frac{\beta C}{\sqrt{2}}\right), \tag{A.6}$$

where $erf(t) := \frac{2}{\sqrt{\pi}} \int_0^t e^{-s^2} ds$. Hence for $C \neq 0$, $(\bar{P}, \bar{m}) = \left(0, -erf\left(\frac{\beta C}{\sqrt{2}}\right)\right)$

(ii) The Jacobian matrix $\left.\frac{\partial(f,g)}{\partial(P,m)}\right|_{\left(0, -erf\left(\frac{\beta C}{\sqrt{2}}\right)\right)}$ has two eigenvalues, one equal to 0 and the other equal to $-\frac{b(1-\bar{m})}{2B + b(1+\bar{m})}$. As the intensity of choice parameter, β, increases from 0 to ∞ the steady state value $\bar{m} = -erf\left(\frac{\beta C}{\sqrt{2}}\right)$ decreases from 0 to -1 which implies that the eigenvalue $-\frac{b(1-\bar{m})}{2B + b(1+\bar{m})}$ decreases from $-\frac{b}{2B+b}$ to $-\frac{b}{B} < -1$. Hence, $\left(0, -erf\left(\frac{\beta C}{\sqrt{2}}\right)\right)$ is a hyperbolic fixed point saddle for some critical value of the switching parameter. For all $\beta < \beta^*$, where $\beta^* = \frac{\sqrt{2} erf^{-1}\left(\frac{B}{b}\right)}{C}$, the eigenvalue $-\frac{b(1-\bar{m})}{2B + b(1+\bar{m})}$ is greater than -1 and hence the system is locally stable. To prove global stability, when

$\beta < \beta^\star$ first note that for all $-1 \leq m_{\rm t} \leq 1, m_{\rm t+1} \geq -erf\left(\frac{\beta C}{\sqrt{2}}\right)$ so we may restrict the analysis to initial states with $m_0 \geq -erf\left(\frac{\beta C}{\sqrt{2}}\right)$. Set $A_{\rm t} = -\frac{b(1-m_{\rm t})}{2B+b(1+m_{\rm t})}$, as we did in part (i), and note that as $m_{\rm t}$ increases from $-erf\left(\frac{\beta C}{\sqrt{2}}\right)$ to 1, then $|A_{\rm t}|$ decreases from $\left|-\frac{b(1-\bar{m})}{2B+b(1+\bar{m})}\right|$ to 0. Hence, $|P_{\rm t+1}| \leq \left|-\frac{b(1-\bar{m}_{\rm t})}{2B+b(1+\bar{m}_{\rm t})}\right|^t P_0$ and since for all $\beta < \beta^\star, \left|-\frac{b(1-\bar{m})}{2B+b(1+\bar{m})}\right| < 1$ we conclude that $P_{\rm t} \to 0$ as $t \to \infty$ which implies that $m_{\rm t} \to 1 - 2\Phi(\beta C)$ or $m_{\rm t} \to -erf\left(\frac{\beta C}{\sqrt{2}}\right)$ using the error function.

(iii) Given than an increase of β leads to a decrease of $-b(1-\bar{m})/(2B+b(1+\bar{m}))$ and for $\beta = \beta^*$, $-b(1-\bar{m})/(2B+b(1+\bar{m})) = -1$, it follows that for $\beta > \beta^*$, $-b(1-\bar{m})/(2B+b(1+\bar{m})) < -1$. Hence as the Jacobian has two eigenvalues one equal to zero and one less than -1, the fixed point is a saddle.

□

Proof of Proposition 3

(i) From the above we see that when $\beta = \beta^*$, the eigenvalue $-b(1-\bar{m})/(2B+b(1+\bar{m})) = -1$, which corresponds to a period doubling bifurcation in which a 2-cycle is created.

The proof of the stability part of the 2-cycle is as follows.

The symmetry with respect to the m-axis implies that the two-cycle must be of the form $\{(P_1, \tilde{m}), (P_2, \tilde{m})\}$ where $P_2 = -\frac{b(1-\tilde{m})}{2B+b(1+\tilde{m})}P_1$. Since the same equality must hold with P_1 and P_2 interchanged we get $-\frac{b(1-\tilde{m})}{2B+b(1+\tilde{m})} = -1$ implying $\tilde{m} = -\frac{B}{b}$. Writing $\tilde{P} = P_1 = -P_2 > 0$ and using equation (22) we find that $\tilde{P}$ satisfies the equation $2\Phi\Big(\beta\Big(2b\tilde{P}^2 - C\Big)\Big) - 1 = -\frac{B}{b}$ or $erf\left(\frac{\beta\left(2b\tilde{P}^2 - C\right)}{\sqrt{2}}\right) = -\frac{B}{b}$ in terms of the error function $erf(t) := \frac{2}{\sqrt{\pi}}\int_0^t e^{-s^2}\,ds$. This equation has a positive solution $\tilde{P} = \sqrt{\frac{C}{2b} + \frac{\sqrt{2}erf^{-1}\left(-\frac{B}{b}\right)}{2b\beta}}$ if and only if $-erf\left(\frac{\beta C}{\sqrt{2}}\right) < -\frac{B}{b}$ which implies that the two-cycle is indeed created when the fixed point becomes unstable at the critical value $\beta^* = \frac{\sqrt{2}erf^{-1}\left(\frac{B}{b}\right)}{C}$. Consider the Taylor series expansion of the dynamics for the states P_t and m_t at the fixed point $(P, m) = (0, \bar{m})$

$$P_{t+1} = \Big[-1 + a_1(m_t - \bar{m}) + a_2(m_t - \bar{m})^2 + ... \Big] P_t + ... \tag{A.7}$$

$$m_{t+1} = \bar{m} + \Big[b_1 + b_2(m_t - \bar{m}) + b_3(m_t - \bar{m})^2 + ... \Big] P_t^2 + ... \tag{A.8}$$

The matrix $\begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix}$ has the following structure: The element -1 corresponds to the coefficient of P_t in equation (35). The element in the first row, second column, corresponds to the coefficient of m_t in the same equation, i.e., equation (35). The two zeros in the second row of the matrix correspond to the coefficients of P_t and m_t in equation (36). Set $x_t = P_t - 0$ and $y_t = m_t - \bar{m}, \forall t$. Then, equations (35) and (36) are written in matrix form as follows

$$\begin{bmatrix} x_{t+1} \\ y_{t+1} \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} + \begin{bmatrix} a_1 y_t x_t + a_2 y_t^2 x_t + ... \\ b_1 x_t^2 + b_2 y_t x_t^2 + b_3 y_t^2 x_t^2 + ... \end{bmatrix}$$

Neglecting terms of degree greater than 2 gives

$$\begin{bmatrix} x_{t+1} \\ y_{t+1} \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} + \begin{bmatrix} a_1 y_t x_t \\ b_1 x_t^2 \end{bmatrix} \tag{A.9}$$

where the coefficients a_1 and b_1 correspond to the values $a_1 = \dfrac{\partial^2 f_1}{\partial P \partial m}\Big(0, \bar{m}\Big)$ and $b_1 = \dfrac{\partial^2 f_2}{\partial P^2}\Big(0, \bar{m}\Big)$. We have used the notation f_1 and f_2 for the functions regarding the dynamics of the states P_t and m_t, respectively, and saved f and g for the purpose of the functional equation

$$h\Big(Ax + f(x, h(x))\Big) - Bh(x) - g\Big(x, h(x)\Big) = 0$$

which is derived in lemma 1 in the online appendix. In this case, $A = -1, B = 0, f(x, y) := a_1 xy$, and $g(x, y) := b_1 x^2$. Hence, the above equation becomes

$$h\Big(-x + a_1 x h(x)\Big) - b_1 x^2 = 0 \tag{A.10}$$

Take h to be $h(x) = c_1x^2 + c_2x^3$. Then, (37) becomes

$$c_1\Big(-x + a_1x(c_1x^2 + c_2x^3)\Big)^2 + c_2\Big(-x + a_1x(c_1x^2 + c_2x^3)\Big)^3 - b_1x^2 = 0 \qquad (A.11)$$

Simplify by keeping only the first and last term in equation (38)

$$c_1\Big(-x + a_1x(c_1x^2 + c_2x^3)\Big)^2 - b_1x^2$$

Solving for c_1 and c_2 gives $c_1 = 2$ and $c_2 = 0$. Hence, $h(x) = 2x^2 + O(x^3)$. This curve is the center manifold of the second-order system (37). Given that $f(x, y) = a_1xy$ it follows that $f\big(x, h(x)\big) = a_1xh(x) = 2a_1x^3$, where

$$a_1 = \frac{\partial^2 f_1}{\partial P \partial m}\big(0, \bar{m}\big) = \frac{b^2(1-\bar{m})}{(2B + b(1+\bar{m}))^2} + \frac{b}{2B + b(a+\bar{m})} > 0.$$

In view of lemma 2, provided in the online appendix, the dynamics on the center manifold is given by the map

$$x_{t+1} = j(x_t) = -x_t + 2a_1x_t^3$$

with a_1 as previously. Hence, by the period doubling bifurcation theorem (e.g. Guckenheimer and Holmes (1983, p. 158)) the two-cycle is stable.

(ii) Let $\underline{F}$ denote the two-dimensional map $\underline{F}(P, m) = \langle f(P, m), g(P, m)\rangle \in \mathbb{R}^2$. The two-cycle is $\Big\{\big(\tilde{P}, \tilde{m}\big), \big(-\tilde{P}, \tilde{m}\big)\Big\}$, where $\tilde{m} = -\frac{B}{b}$ and $\tilde{P}$ is the positive solution of the equation $\frac{1}{2}\Big\{erf\Big[\frac{\beta}{\sqrt{2}}\Big(2b\tilde{P}^2 - C\Big)\Big] - erf\Big[\frac{\beta}{\sqrt{2}}\Big(C - 2b\tilde{P}^2\Big)\Big]\Big\} = erf\Big[\frac{\beta}{\sqrt{2}}\Big(2b\tilde{P}^2 - C\Big)\Big] = \tilde{m} = -\frac{B}{b}$. Using $f\big(\tilde{P}, \tilde{m}\big) = -\tilde{P}$, which implies $f_P\big(\pm\tilde{P}, \tilde{m}\big) = -1$, a straightforward computation yields the Jacobian matrix at the period 2 point $\big(\tilde{P}, \tilde{m}\big)$

$$\frac{\partial(f, g)}{\partial(P, m)}\bigg|_{\big(\tilde{P}, \tilde{m}\big)} = \begin{bmatrix} -1 & \frac{2b\tilde{P}}{b+B} \\ 2\sqrt{2}b\beta\tilde{P}erf'\Big(\frac{\beta}{\sqrt{2}}\Big(2b\tilde{P}^2 - C\Big)\Big) & -2\sqrt{2}b\beta\frac{b}{b+B}\tilde{P}^2erf'\Big(\frac{\beta}{\sqrt{2}}\Big(2b\tilde{P}^2 - C\Big)\Big) \end{bmatrix} \qquad (40)$$

or writing $\gamma = \frac{b\tilde{P}}{b+B}$ and $\delta = 2\sqrt{2}b\beta\tilde{P}erf'\Big(\frac{\beta}{\sqrt{2}}(2b\tilde{P}^2 - C)\Big)$,

$$\frac{\partial(f, g)}{\partial(P, m)}\bigg|_{\big(\tilde{P}, \tilde{m}\big)} = \begin{bmatrix} -1 & 2\gamma \\ \delta & -\gamma\delta \end{bmatrix} \text{ and } \frac{\partial(f, g)}{\partial(P, m)}\bigg|_{\big(-\tilde{P}, \tilde{m}\big)} = \begin{bmatrix} -1 & -2\gamma \\ -\delta & -\gamma\delta \end{bmatrix} \qquad (41)$$

The Jacobian matrix of the second iterate, $\underline{F}^2(P,m)$, at $\Big(\tilde{P},\tilde{m}\Big)$, is given by the matrix product

$$\left.\frac{\partial(f,g)}{\partial(P,m)}\right|_{\Big(-\tilde{P},\tilde{m}\Big)} \cdot \left.\frac{\partial(f,g)}{\partial\Big(P,m\Big)}\right|_{\Big(\tilde{P},\tilde{m}\Big)} = \begin{bmatrix} 1-2\gamma\delta & 2\gamma^2\delta - 2\gamma \\ \delta - \gamma\delta^2 & -2\gamma\delta + \gamma^2\delta^2 \end{bmatrix} \tag{42}$$

The characteristic equation of the matrix in (42) is

$$s^2 + (4\gamma\delta - \gamma^2\delta^2 - 1)s + \gamma^2\delta^2 = 0 \tag{43}$$

with discriminant $D = (\gamma\delta - 1)^2(\gamma^2\delta^2 - 6\gamma\delta + 1)$ which is strictly negative for $\gamma\delta \in (3-2\sqrt{2},1) \cup (1,3+2\sqrt{2})$. As $\beta \longrightarrow \infty$ it holds that $\tilde{P} \longrightarrow \sqrt{\frac{C}{2b}}$, so $\frac{b\tilde{P}}{b+B} \longrightarrow \frac{b}{b+B}\sqrt{\frac{C}{2b}}$. Furthermore, it holds that $erf'\Big[\frac{\beta}{\sqrt{2}}\Big(2b\tilde{P}^2 - C\Big)\Big] = erf'\Big(erf^{-1}(-\frac{B}{b})\Big)$. Hence, $\gamma\delta \longrightarrow \infty$ as $\beta \longrightarrow \infty$. Let β^{**} and β_1 be the values of β for which $\gamma\delta = 1$ and $\gamma\delta = 3+2\sqrt{2}$, respectively. At $\beta = \beta^{**}$ the Jacobian matrix in (42) becomes $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$ wich has a pair of eigenvalues $s_1 = s_2 = -1$ with modulus 1. Hence, at $\beta = \beta^{**}$, the eigenvalues cross the unit disk, from the inside to the outside, a clear sign of the appearance of a Hopf bifurcation. It follows that for $\beta < \beta^{**}$, but sufficiently close to β^{**} such that $\gamma\delta$ is in the interval $\Big(3-2\sqrt{2},1\Big)$, the two-cycle $\Big\{\Big(\tilde{P},\tilde{m}\Big),\Big(-\tilde{P},\tilde{m}\Big)\Big\}$ is stable with complex eigenvalues. For $\beta^{**} < \beta$, with β sufficiently close to β^{**} such that $\gamma\delta$ is in the interval $\Big(1,3+2\sqrt{2}\Big)$, the two-cycle $\Big\{\Big(\tilde{P},\tilde{m}\Big),\Big(-\tilde{P},\tilde{m}\Big)\Big\}$ is unstable with complex eigenvalues since the discriminant $D < 0$.

□

B Estimated results for two choice alternatives

Table 3: Maximum likelihood estimation for experimental data with two choice alternatives

	SI2			BH2		
	logit	probit	probit-RT	logit	probit	probit-RT
β	3.747	2.161	2.161	0.431	0.255	0.288
	(0.277)	(0.141)	(0.141)	(0.034)	(0.019)	(0.034)
σ_b			0.000			0.125
						(0.062)
log-likelihood	-203.347	-202.950	-202.950	-285.560	-286.284	-285.155
AIC	408.694	**407.899**	409.900	**573.121**	574.567	574.310
BIC	413.022	**412.227**	418.555	**577.484**	578.930	583.036

As in Anufriev et al. (2016), estimations are based on the final 20 experimental periods. SI2 is based on stock market time series and BH2 is based on Brock and Hommes (1997). The estimations of the columns labeled 'logit' and 'probit' are discussed in Section 3.3. The estimations of the columns labeled 'probit-RT' (random taste variation) are discussed in Section 4.

Online Appendix (Not for publication)

C Derivations and Lemmas

Derivation of Equation (19)

We have

$$n_{P,t+1} = \Phi\left(\beta\left[\frac{b}{2}\left(P_{t+1}-P_t\right)^2 - C\right]\right),$$

$$n_{N,t+1} = \Phi\left(\beta\left[C - \frac{b}{2}\left(P_{t+1}-P_t\right)^2\right]\right).$$

Then the difference in fractions is

$$m_{t+1} = n_{P,t+1} - n_{N,t+1} =$$

$$= \Phi\left(\beta\left[\frac{b}{2}\left(P_{t+1} - P_t\right)^2 - C\right]\right) - \Phi\left(\beta\left[C - \frac{b}{2}\left(P_{t+1} - P_t\right)^2\right]\right).$$

Plugging (18) to the above we get

$$m_{t+1} = n_{P,t+1} - n_{N,t+1} =$$

$$= \Phi\left(\beta\left[\frac{b}{2}\left(\frac{b(1-m_t)P_t}{2B + b(1+m_t)} + P_t\right)^2 - C\right]\right) - \Phi\left(\beta\left[C - \frac{b}{2}\left(\frac{b(1-m_t)P_t}{2B + b(1+m_t)} + P_t\right)^2\right]\right).$$

Since $\Phi(-z) = 1 - \Phi(z)$, $-\infty < z < \infty$, the previous can be written equivalently as

$$m_{t+1} = 2\Phi\Big(\beta\Big[\frac{b}{2}\Big(\frac{b(1-m_t)P_t}{2B + b(1+m_t)} + P_t\Big)^2 - C\Big]\Big) - 1.$$

Lemmas supporting the proof of Proposition 3

Lemma 1. *Consider a sufficiently smooth two-dimensional map $\underline{F}_\beta$ where β is a bifurcation parameter. Let $(\overline{x}, \overline{y})$ be a fixed point of $\underline{F}_\beta$, for $\beta = \beta_0$, which we translate, without loss of generality, to the origin and write the system in a Taylor series expansion as $x_{t+1} = Ax_t + f(x_t, y_t), y_{t+1} = By_t + g(x_t, y_t)$. Then, the center manifold curve $y_t = h(x_t)$ of the map $x_{t+1} = Ax_t + f(x_t, y_t), y_{t+1} = By_t + g(x_t, y_t)$ satisfies the functional equation $h[Ax_t + f(x_t, h(x_t))] - Bh(x_t) - g(x_t, h(x_t)) = 0$*

Proof

Indeed, $y_t = h(x_t)$ implies $h[Ax_t + f(x_t, h(x_t))] - Bh(x_t) - g(x_t, h(x_t)) = 0$. This is a functional equation where the unknown is the function $y_t = h(x_t)$, i.e., the center manifold curve.

□

Lemma 2. *Let $\underline{F}_\beta$ be a sufficiently smooth two-dimensional map which has a fixed point, $(\overline{x}, \overline{y})$, translated, without loss of generality, to the origin. For a fixed value, β^*, of the*

bifurcation parameter, β, the dynamics of the second-order system $x_{t+1} = Ax_t + f(x_t, y_t)$, $y_{t+1} = By_t + g(x_t, y_t)$, where $Df(0,0) = Dg(0,0) = 0$, restricted on the center manifold are given locally by the map $x_{t+1} = Ax_t + f(x_t, h(x_t))$.

Proof

The change of variables $x \to x, w \to y - h(x)$ transforms the map $x \to Ax + f(x,y), y \to By+g(x,y)$, corresponding to the second-order difference equation $x_{t+1} = Ax_t+f(x_t, y_t), y_{t+1} = By_t + g(x_t, y_t)$, into

$$x_{t+1} = Ax_t + f\Big(x_t, w_t + h(x_t)\Big) \tag{C.1}$$

$$w_{t+1} = B\Big(w_t + h(x_t)\Big) + g\Big(x_t, w_t + h(x_t)\Big) - h\Big(Ax_t + f(x_t, w_t + h(x_t))\Big) \tag{C.2}$$

In the new coordinates, the dynamics on the center manifold is characterized by $w_t \equiv 0 \Leftrightarrow \Delta w_t \equiv 0$. Substituting this identity into equation (33) yields

$$0 = Bh(x_t) + g\Big(x_t, h(x_t)\Big) - h\Big(Ax_t + f(x_t, h(x_t))\Big) \tag{C.3}$$

Adding and subtracting $f\Big(x_t, h(x_t)\Big)$ in the right-hand side of equation (32) and also subtracting equation (34) from equation (33) yields the transformed dynamics

$$x_{t+1} = Ax_t + f\Big(x_t, h(x_t)\Big) + N(x_t, w_t)$$

$$w_{t+1} = Bw_t + R(x_t, w_t)$$

in the new coordinates x, w , where

$N(x_t, w_t) := f\Big(x_t, w_t + h(x_t)\Big) - f\Big(x_t, h(x_t)\Big)$

and

$R(x_t, w_t) = g\Big(x_t, w_t + h(x_t)\Big) - h\Big(Ax_t + f(x_t, w_t + h(x_t))\Big) - g\Big(x_t, h(x_t)\Big) + h\Big(Ax_t + f(x_t, h(x_t))\Big)$. It is easily verified that $N(x,0) = R(x,0) = \dfrac{\partial N}{\partial w}(0,0) = \dfrac{\partial R}{\partial w}(0,0) = 0$. Hence, in the domain $|x| < \rho, |w| < \rho$ it holds that $|N(x,w)| \leq k_1|w|$ and $|R(x,w)| \leq k_2|w|$, where the constants k_1 and k_2 can be made arbitrarily small by choosing ρ sufficiently small. In addition, given that $|B| < 1$, the stability properties of the origin are determined by the reduced system $x \to Ax + f\Big(x, h(x)\Big)$ □

D Estimation on Experimental Data

D.1 Derivation of variance covariance matrices

The general form of the variance-covariance matrix for the three-alternative case is given by the square, symmetric matrix:

$$\Omega = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix} \tag{D.1}$$

The variance-covariance matrix of error differences for the first alternative is given by the product:

$$\tilde{\Omega}_A = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix} \begin{bmatrix} -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{D.2}$$

Carrying out this multiplication yields

$$\tilde{\Omega}_A = \begin{bmatrix} \sigma_{11} + \sigma_{22} - 2\sigma_{12} & \sigma_{11} + \sigma_{23} - \sigma_{12} - \sigma_{13} \\ \sigma_{11} + \sigma_{23} - \sigma_{12} - \sigma_{13} & \sigma_{11} + \sigma_{33} - 2\sigma_{13} \end{bmatrix} \tag{D.3}$$

Setting

$$\begin{bmatrix} \theta_{22} \\ \theta_{23} \\ \theta_{33} \end{bmatrix} = \begin{bmatrix} 1 & -2 & 0 & 1 & 0 & 0 \\ 1 & -1 & -1 & 0 & 1 & 0 \\ 1 & 0 & -2 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{11} \\ \sigma_{12} \\ \sigma_{13} \\ \sigma_{22} \\ \sigma_{23} \\ \sigma_{33} \end{bmatrix}, \tag{D.4}$$

(D.3) becomes

$$\tilde{\Omega}_A = \begin{bmatrix} \theta_{22} & \theta_{23} \\ \theta_{23} & \theta_{33} \end{bmatrix} \tag{D.5}$$

which can be normalized to:

$$\tilde{\Omega}_A^* = \begin{bmatrix} 1 & \theta_{23}^* \\ \theta_{23}^* & \theta_{33}^* \end{bmatrix}, \tag{D.6}$$

where

$$\theta_{23}^* = \frac{\theta_{23}}{\theta_{22}} = \frac{\sigma_{11} - \sigma_{12} - \sigma_{13} + \sigma_{23}}{\sigma_{11} - 2\sigma_{12} + \sigma_{22}}$$

$$\theta_{33}^* = \frac{\theta_{33}}{\theta_{22}} = \frac{\sigma_{11} - 2\sigma_{13} + \sigma_{33}}{\sigma_{11} - 2\sigma_{12} + \sigma_{22}}$$

The variance-covariance matrix is positive definite. Then, by the Cholesky analysis theorem, $\tilde{\Omega}_A^*$ can be analyzed in a unique way as

$$\tilde{\Omega}_A^* = L_A L_A' \tag{D.7}$$

where L_A is a lower triangular matrix whose principal diagonal elements are strictly positive. L_A is called the Cholesky factor of $\tilde{\Omega}_A^*$. Here:

$$\tilde{\Omega}_A^* = \begin{bmatrix} 1 & 0 \\ \theta_{23}^* & \sqrt{\theta_{33}^* - (\theta_{23}^*)^2} \end{bmatrix} \begin{bmatrix} 1 & \theta_{23}^* \\ 0 & \sqrt{\theta_{33}^* - (\theta_{23}^*)^2} \end{bmatrix} \tag{D.8}$$

The above shows that we have two free parameters. For the maximum likelihood estimation, we proceed by parameterizing the two unknown elements (ϕ_{21} and ϕ_{22}) of

$$L_A = \begin{bmatrix} 1 & 0 \\ \phi_{21} & \phi_{22} \end{bmatrix} \tag{D.9}$$

as parameters to be estimated. From there, we calculate $\tilde{\Omega}_A^*$ using (D.7). Or in other words, we have that

$$\theta_{23}^* = \phi_{21} \tag{D.10}$$

$$\theta_{33}^* = (\phi_{21})^2 + (\phi_{22})^2 \tag{D.11}$$

Next, we use similar steps as above to express the (normalized) variance-covariance matrix of error differences for the second alternative in terms of θ_{23}^* and θ_{33}^* so that this

matrix is also pinned down by the estimated values of ϕ_{21} and ϕ_{22}.

In particular, we first use that

$$\tilde{\Omega}_B = \begin{bmatrix} 1 & -1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & -1 \\ 0 & 1 \end{bmatrix} \tag{D.12}$$

which reduces to

$$\tilde{\Omega}_B = \begin{bmatrix} \sigma_{22} + \sigma_{11} - 2\sigma_{12} & \sigma_{13} + \sigma_{22} - \sigma_{12} - \sigma_{23} \\ \sigma_{13} + \sigma_{22} - \sigma_{12} - \sigma_{23} & \sigma_{22} + \sigma_{33} - 2\sigma_{23} \end{bmatrix} \tag{D.13}$$

Using (D.4), this can be written as

$$\tilde{\Omega}_B = \begin{bmatrix} \theta_{22} & \theta_{22} - \theta_{23} \\ \theta_{22} - \theta_{23} & \theta_{22} + \theta_{33} - 2\theta_{23} \end{bmatrix} \tag{D.14}$$

which, after normalization, can then be expressed in terms of the normalized coefficients θ^*_{23} and θ^*_{33} as

$$\tilde{\Omega}^*_B = \begin{bmatrix} 1 & 1 - \theta^*_{23} \\ 1 - \theta^*_{23} & 1 + \theta^*_{33} - 2\theta^*_{23} \end{bmatrix} \tag{D.15}$$

Using the resulting $\tilde{\Omega}^*_A$ and $\tilde{\Omega}^*_B$, we calculate the fractions of agents following the first two alternatives using a numerical implementation of the CDF of a multinomial probit distribution in MATLAB. The fraction of agents following the third alternative is then pinned down by the fact that the fractions of agents following the three alternatives must sum up to 1.

For the case of four choice alternatives we proceed in the same way. Consider initially

the product

$$\tilde{\Omega}_A = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_{44} \end{bmatrix} \begin{bmatrix} -1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \iff$$

$$\iff \tilde{\Omega}_A = \begin{bmatrix} \sigma_{11} + \sigma_{22} - 2\sigma_{12} & \sigma_{11} + \sigma_{23} - \sigma_{12} - \sigma_{13} & \sigma_{11} + \sigma_{24} - \sigma_{12} - \sigma_{14} \\ \sigma_{11} + \sigma_{23} - \sigma_{12} - \sigma_{13} & \sigma_{11} + \sigma_{33} - 2\sigma_{13} & \sigma_{11} + \sigma_{34} - \sigma_{13} - \sigma_{14} \\ \sigma_{11} + \sigma_{24} - \sigma_{12} - \sigma_{14} & \sigma_{11} + \sigma_{34} - \sigma_{13} - \sigma_{14} & \sigma_{11} + \sigma_{44} - 2\sigma_{14} \end{bmatrix} \tag{D.16}$$

Next, we define

$$\begin{bmatrix} \theta_{22} \\ \theta_{33} \\ \theta_{44} \\ \theta_{23} \\ \theta_{24} \\ \theta_{34} \end{bmatrix} = \begin{bmatrix} 1 & -2 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -2 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & -1 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \sigma_{11} \\ \sigma_{12} \\ \sigma_{13} \\ \sigma_{14} \\ \sigma_{22} \\ \sigma_{23} \\ \sigma_{24} \\ \sigma_{33} \\ \sigma_{34} \\ \sigma_{44} \end{bmatrix} \tag{D.17}$$

to write (D.16) as

$$\tilde{\Omega}_A = \begin{bmatrix} \theta_{22} & \theta_{23} & \theta_{24} \\ \theta_{23} & \theta_{33} & \theta_{34} \\ \theta_{24} & \theta_{34} & \theta_{44} \end{bmatrix} \tag{D.18}$$

which can be normalized to:

$$\tilde{\Omega}_A^* = \begin{bmatrix} 1 & \theta_{23}^* & \theta_{24}^* \\ \theta_{23}^* & \theta_{33}^* & \theta_{34}^* \\ \theta_{24}^* & \theta_{34}^* & \theta_{44}^* \end{bmatrix} \tag{D.19}$$

where:

$$\theta_{23}^* = \frac{\theta_{23}}{\theta_{22}} = \frac{\sigma_{11} + \sigma_{23} - \sigma_{12} - \sigma_{13}}{\sigma_{11} + \sigma_{22} - 2\sigma_{12}},$$

$$\theta_{24}^* = \frac{\theta_{24}}{\theta_{22}} = \frac{\sigma_{11} + \sigma_{24} - \sigma_{12} - \sigma_{14}}{\sigma_{11} + \sigma_{22} - 2\sigma_{12}},$$

$$\theta_{33}^* = \frac{\theta_{33}}{\theta_{22}} = \frac{\sigma_{11} + \sigma_{33} - 2\sigma_{13}}{\sigma_{11} + \sigma_{22} - 2\sigma_{12}},$$

$$\theta_{34}^* = \frac{\theta_{34}}{\theta_{22}} = \frac{\sigma_{11} + \sigma_{34} - \sigma_{13} - \sigma_{14}}{\sigma_{11} + \sigma_{22} - 2\sigma_{12}},$$

$$\theta_{44}^* = \frac{\theta_{44}}{\theta_{22}} = \frac{\sigma_{11} + \sigma_{44} - 2\sigma_{14}}{\sigma_{11} + \sigma_{22} - 2\sigma_{12}}.$$

Finally, one can verify that the Cholesky decomposition of the matrix $\tilde{\Omega}_A^*$ is the following:

$$\begin{bmatrix} 1 & 0 & 0 \\ \theta_{23}^* & \sqrt{\theta_{33}^* - (\theta_{23}^*)^2} & 0 \\ \theta_{24}^* & \frac{\theta_{34}^* - \theta_{23}^*\theta_{24}^*}{\sqrt{\theta_{33}^* - (\theta_{23}^*)^2}} & \sqrt{\theta_{44}^* - (\theta_{24}^*)^2 - \frac{(\theta_{34}^* - \theta_{23}^*\theta_{24}^*)^2}{\theta_{33}^* - (\theta_{23}^*)^2}} \end{bmatrix} \begin{bmatrix} 1 & \theta_{23}^* & \theta_{24}^* \\ 0 & \sqrt{\theta_{33}^* - (\theta_{23}^*)^2} & \frac{\theta_{34}^* - \theta_{23}^*\theta_{24}^*}{\sqrt{\theta_{33}^* - (\theta_{23}^*)^2}} \\ 0 & 0 & \sqrt{\theta_{44}^* - (\theta_{24}^*)^2 - \frac{(\theta_{34}^* - \theta_{23}^*\theta_{24}^*)^2}{\theta_{33}^* - (\theta_{23}^*)^2}} \end{bmatrix} \tag{D.20}$$

For the maximum likelihood estimation we use the following paramterization of the Cholesky factor

$$L_A = \begin{bmatrix} 1 & 0 & 0 \\ \phi_{21} & \phi_{22} & 0 \\ \phi_{31} & \phi_{32} & \phi_{33} \end{bmatrix} \tag{D.21}$$

so that

$$\theta_{23}^* = \phi_{21} \tag{D.22}$$

$$\theta_{33}^* = (\phi_{21})^2 + (\phi_{22})^2 \tag{D.23}$$

$$\theta_{24}^* = \phi_{31} \tag{D.24}$$

$$\theta^*_{34} = \phi_{21}\phi_{31} + \phi_{22}\phi_{32} \tag{D.25}$$

$$\theta^*_{44} = (\phi_{31})^2 + (\phi_{32})^2 + (\phi_{33})^2 \tag{D.26}$$

Analogue to the above, we derive

$$\tilde{\Omega}^*_B = \begin{bmatrix} 1 & 1-\theta^*_{23} & 1-\theta^*_{24} \\ 1-\theta^*_{23} & 1+\theta^*_{33}-2\theta^*_{23} & 1+\theta^*_{34}-\theta^*_{23}-\theta^*_{24} \\ 1-\theta^*_{24} & 1+\theta^*_{34}-\theta^*_{23}-\theta^*_{24} & 1+\theta^*_{44}-2\theta^*_{24} \end{bmatrix} \tag{D.27}$$

and

$$\tilde{\Omega}^*_C = \begin{bmatrix} \theta^*_{33} & \theta^*_{33}-\theta^*_{23} & \theta^*_{33}-\theta^*_{34} \\ \theta^*_{33}-\theta^*_{23} & 1+\theta^*_{33}-2\theta^*_{23} & \theta^*_{33}+\theta^*_{24}-\theta^*_{23}-\theta^*_{34} \\ \theta^*_{33}-\theta^*_{34} & \theta^*_{33}+\theta^*_{24}-\theta^*_{23}-\theta^*_{34} & \theta^*_{33}+\theta^*_{44}-2\theta^*_{34} \end{bmatrix} \tag{D.28}$$

D.2 Estimated variances covariance matrices of error differences

For SI3, the estimated ϕ coefficients imply

$$\tilde{\Omega}^*_A = \begin{pmatrix} 1.000 & 0.867 \\ 0.867 & 1.181 \end{pmatrix}$$

$$\tilde{\Omega}^*_B = \begin{pmatrix} 1.000 & 0.133 \\ 0.133 & 0.446 \end{pmatrix}$$

For BH3, we have

$$\tilde{\Omega}^*_A = \begin{pmatrix} 1.000 & 0.582 \\ 0.582 & 0.839 \end{pmatrix}$$

$$\tilde{\Omega}^*_B = \begin{pmatrix} 1.000 & 0.418 \\ 0.418 & 0.675 \end{pmatrix}$$

For the case of four choice alternatives we first have for SI4

$$\tilde{\Omega}^*_A = \begin{pmatrix} 1.000 & 1.079 & 1.015 \\ 1.079 & 1.507 & 1.387 \\ 1.015 & 1.387 & 1.354 \end{pmatrix}$$

$$\tilde{\Omega}^*_B = \begin{pmatrix} 1.000 & -0.079 & -0.015 \\ -0.079 & 0.349 & 0.293 \\ -0.015 & 0.293 & 0.324 \end{pmatrix}$$

$$\tilde{\Omega}^*_C = \begin{pmatrix} 1.507 & 0.428 & 0.120 \\ 0.428 & 0.349 & 0.056 \\ 0.120 & 0.056 & 0.087 \end{pmatrix}$$

And further for BH4

$$\tilde{\Omega}^*_A = \begin{pmatrix} 1.000 & 0.999 & 0.999 \\ 0.999 & 1.010 & 1.004 \\ 0.999 & 1.004 & 1.002 \end{pmatrix}$$

$$\tilde{\Omega}^*_B = \begin{pmatrix} 1.000 & 0.001 & 0.001 \\ 0.001 & 0.012 & 0.006 \\ 0.001 & 0.006 & 0.003 \end{pmatrix}$$

$$\tilde{\Omega}^*_C = \begin{pmatrix} 1.010 & 0.011 & 0.006 \\ 0.011 & 0.012 & 0.006 \\ 0.006 & 0.006 & 0.003 \end{pmatrix}$$