

# UC Berkeley

## UC Berkeley Electronic Theses and Dissertations

### Title

Essays in Behavioral Economics and Risk Management

### Permalink

<https://escholarship.org/uc/item/1x23v47p>

### Author

HOLMAN, JEFFREY TREVOR

### Publication Date

2010

Peer reviewed|Thesis/dissertation

Essays in Behavioral Economics and Risk Management

By

Jeffrey Trevor Holman

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Robert M. Anderson, Chair

Professor Stefano DellaVigna

Professor Dwight Jaffee

Fall 2010

The dissertation of Jeffrey Trevor Holman, titled Essays in Behavioral Economics and Risk Management, is approved:

Chair \_\_\_\_\_ Date \_\_\_\_\_  
Robert M. Anderson

\_\_\_\_\_ Date \_\_\_\_\_  
Stefano DellaVigna

\_\_\_\_\_ Date \_\_\_\_\_  
Dwight Jaffee

University of California, Berkeley

## Abstract

Essays in Behavioral Economics and Risk Management

by

Jeffrey Trevor Holman

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Robert M. Anderson, Chair

In the first chapter, I develop a model of prospective memory, defined as the capacity to recall actions to be carried out in the future. An agent faces some task with stochastic cost  $c_t$ , benefit  $b$ , and  $T$  periods until some exogenously imposed deadline. The agent can only execute the task at time  $t$  if the task is recalled in that period. The memory process exhibits the rehearsal property that the probability of recall is lower if the task was forgotten in the recent past. The agent sets a threshold cost each period based on her expectations of whether she will recall and carry out the task in future periods. If the task is recalled at time  $t$ , and the draw from the cost distribution is below this threshold, the task is executed. We then introduce memory overconfidence into the model, which we define as either overestimating the base likelihood of recall in future periods or underestimating the effect of temporary forgetting on subsequent recall. Memory overconfidence leads not only to inefficiently low rates of task completion, but also to the prediction that the probability of task completion may vary inversely with the length of time allocated to completing the task. We discuss the interaction of these effects with present-biased preferences, and provide examples of economic scenarios where this dynamic may be exploited by firms to the detriment of consumers.

In the second chapter, I introduce a new copula which simultaneously allows fully-general correlation structures in the bulk of a multivariate distribution and an arbitrarily high degree of dependence in the left tails. This is ideally suited for modeling financial assets which may display moderate cor-

relation in normal times, but which experience simultaneous left tail events, such as during a financial crisis. The new copula is shown to be fully flexible in the sense that the user can specify a desired structure for a sequence of increasingly dire events in the left tail, while still retaining the same correlation structure in the bulk. Finally, I illustrate the use of this copula with an application to hedge fund returns.

# 1 Prospective Memory

## 1.1 Introduction

Human memory is far from perfect. Beginning with Ebbinghaus’s groundbreaking work on the “forgetting curve” in 1885 [21], the psychology literature contains extensive evidence supporting this claim. Yet, in economics, the critical role that fallible memory plays in coloring individual decision-making has been largely neglected until recently. This neglect may lead to the misattribution of certain observed anomalies to other types of biases.

One such anomaly is the failure of individuals to carry out projects with small, immediate costs and large, deferred benefits. Several empirical studies have documented this finding and explained it by time-inconsistent preferences.<sup>1</sup> For instance, DellaVigna and Malmendier [20] suggest that the low cancelation rates observed with automatically renewed health club memberships may be an example of status quo bias generated by hyperbolic time-discounting. Madrian and Shea [54] and Choi, Laibson, Madrian, and Metrick [15] suggest that this same status quo bias keeps employees from updating the often-suboptimal default enrollment options in 401(k) plans.

While we agree that time-inconsistency plays a role in generating this inefficient behavior, we believe that this explanation is incomplete, particularly when considering projects with deadlines. While present-biased preferences may lead to delay in executing beneficial tasks such as canceling an unused health club membership or modifying a benefits plan, such preferences often cannot explain the *failure* to carry out such tasks altogether. For example, consider an individual who has 30 days to mail in a \$50 consumer rebate. It is calibrationally feasible to derive from present-biased preferences that, for the first 29 days, she will defer completing and mailing out the rebate form, preferring to do it “tomorrow” over “today.” However, on the 30th day, her choice is between “today” and “never”, and justifying task omission in such cases often requires assuming an extremely low  $\beta$  (i.e., extreme present-biased

---

<sup>1</sup>In the quasi-hyperbolic form, time-inconsistent agents have a discount rate of  $\beta\delta$  between the current period and the next period, and  $\delta$  between all future periods. Thus,  $\beta$  captures the special salience of the present.

preferences).

Assuming stochastic costs and naivete with respect to present-biased preferences, as in O’Donoghue and Rabin [62], can lead to inefficiently low task execution rates, even for feasible values of  $\beta$ . However, these assumptions are inconsistent with the empirical finding that task execution rates may vary inversely with the length of deadline. Shafir and Tversky [75] offered students \$5 to return a long questionnaire by a given date. The students were randomly assigned to one of three deadline groups – the first group was given 5 days, the second group 3 weeks, and the third group no definite deadline. The respective rates of return for the three groups were 60%, 42%, and 25%. Procrastination can explain the drop in the return rate between the groups with a deadline and the group with no deadline – when no deadline is given, students never reach the “now” or “never” decision, and naive hyperbolics will defer completing the survey each period, always planning on doing it next period. However, with a finite deadline, a model with stochastic costs and naive hyperbolic time-discounting would predict a *positive* relationship between the likelihood of task execution and the length of deadline. Having more time gives agents more chances to draw a low-cost realization, and thus the likelihood of executing the task would increase.

Perhaps most importantly, the time-inconsistency explanation simply feels incomplete when we relate it to our personal experiences in failing to execute basic, time-sensitive tasks. Consider the last time you failed to mail in a rebate, cancel a free trial offer, or return a borrowed movie or book on time. What happened? A common response is simply, “I forgot.” In this paper, we explore this type of fallible memory as an alternative (or more accurately, complement) to time-inconsistent preferences in explaining inefficient behavior with respect to completing tasks with deadlines.

We build on the work of Mullainathan [60], which developed a retrospective memory-based model of bounded rationality, by extending the analysis to prospective memory. The distinction, as explained by psychologist P.E. Morris (1992), is that “prospective memory is memory for intentions, for actions that we wish to carry out in the future, while retrospective memory is recall of information from the past.” There is continued debate among psychologists on whether a distinction should be made between these two “types” of memory, given that the underlying mechanisms that determine the success or failure of

remembering a past memory or a future task may be the same. Nonetheless, from the perspective of economics, we view these as distinct types of memory recall, with different implications for individual decision-making behavior.

Section 1.2 surveys the existing psychology literature on fallible memory, and prospective recall in particular. The experimental evidence on prospective memory is quite limited. The most common studies – “postcard studies” – instruct subjects to mail in postcards at (or before) some specified future date, and monitor how return rates vary with time frame, cues, and monetary incentives. Unfortunately, many of these studies have limited sample sizes and other methodological concerns, and are uninformative on how prospective memory beliefs correlate with actual recall performance. More recent work in the economics and marketing literature ([77] and [27]) addressed some of these concerns.

In Section 1.3, we develop a baseline model of prospective memory (PM), and apply it to a choice problem where the agent faces some task with stochastic cost  $c_t$ , fixed benefit  $b$ , and  $T$  periods until some exogenously imposed deadline. The agent can only execute the task at time  $t$  if the task is recalled in that period. We adopt a memory function with the rehearsal property described in Mullainathan [60] – that is, the probability of recall is lower if the event (or in this case, the task) was forgotten in the recent past. The agent sets a threshold cost each period based on her expectations of whether she will recall and carry out the task in future periods. If the task is recalled at time  $t$ , and the draw from the cost distribution is below this threshold, the task is executed. We discuss the time pattern of hazard rates for agents with perfect and imperfect memories, and demonstrate an empirically testable prediction for fallible memory – hazard rates must be increasing for agents with perfect memory, but may be decreasing for agents with imperfect memory.

In our baseline model, we assume that agents, despite their fallible memory, have correct beliefs over their PM process. In Section 1.4, we introduce the possibility that agents have systematically incorrect PM beliefs, as suggested by experimental evidence in Silk [77] and Ericson [27] on subjects’ incentive-compatible memory beliefs and subsequent performance. In particular, agents may be overconfident in their prospective memory. We define prospective memory overconfidence as overestimating the base likelihood of recall in future periods, or underestimating the effect of forgetting the task in



any period (or series of periods) on the subsequent probability of recall. We view PM overconfidence as a form of projection bias [51]. That is, individuals “project” their current memory state onto all future periods, and develop their expectations for future recall and behavior accordingly. PM overconfidence can also be viewed as analogous to other forms of “information projection” such as hindsight bias and curse of knowledge.

In Section 1.5, we discuss the effects of PM overconfidence on agent behavior in our model. First, PM overconfidence reduces welfare. When an agent is overconfident with respect to prospective memory, the perceived continuation value of deferring tasks is higher than the true continuation value – the agent will inefficiently defer tasks, overoptimistically relying on future recall and execution. Second, in our key result, PM overconfidence increases the likelihood that extending the deadline will be to the detriment of the agent – that is, that both the ex ante expected utility and probability of task execution decrease as  $T$  increases. While the optimal strategy requires agents to decrease threshold costs when the deadline is extended, overconfident agents, in overestimating the benefit of the extended deadline, reduce threshold costs by more than they should. Under certain conditions (in particular, if agents are sufficiently forgetful and overconfident), the cost of this over-selectivity in early periods outweighs the benefit of having more periods to remember the task, and agents are made worse off. Thus, our model provides an explanation for the Shafir and Tversky student survey result.

In Section 1.6 we discuss two extensions of the model – memory aids and time discounting. We incorporate memory aids by assuming that, for some cost, agents can ensure recall in some subset of the periods before the deadline. Depending on the values of the memory parameters, agents will either choose to backload reminders (to allow for lower threshold costs in early periods) or spread out reminders (to mitigate the adverse effect of successive forgetting). We also show that, for an increasing benefit from task completion, the welfare loss from fallible memory for overconfident agents can be made arbitrarily large, while for agents with correct PM beliefs, the welfare loss is bounded by the cost of a complete memory aid.

We next incorporate time-discounting into the model, and demonstrate that the Shafir and Tversky result cannot be explained by hyperbolic discounting alone. At the same time, we show that time-inconsistency has positive

interaction effects with PM overconfidence – procrastination further encourages agents to defer tasks, making their overestimate of the continuation value of deferral more costly. Indeed, we believe that individuals’ failure to execute basic, time-sensitive tasks in the real world is not due solely to imperfect memory or present-biased preferences, but rather some combination of the two.

Rebates and free trial offers are two “real world” mechanisms by which firms appear to be exploiting naive forgetting by consumers. In Section 1.7, we provide an overview of these prospective memory failures in the marketplace. We present anecdotal evidence on the high degree of ex post consumer regret and frustration over rebates and trial offers, and discuss how this evidence runs counter to the rational-model explanations for these marketing tactics. We believe these areas are ripe for potential field studies identifying and calibrating the effects of prospective memory failures and overconfidence (Section 1.8). We conclude with some thoughts on potential theoretical extensions of our model.

## **1.2 The Psychology Literature on Retrospective and Prospective Memory**

### **1.2.1 Forgetting and The Retention Function**

Much of the psychology literature on memory focused on retrospective recall. Ebbinghaus [21] is widely recognized as the first experimental study to attempt to measure the rate of forgetting, and is credited for providing the first empirical evidence that memory retention is nonlinear in time.<sup>2</sup>

Since then, the memory retention function has been estimated in experiments of various contexts. Studies have varied the time frame (from minutes to years) and the content of information to be remembered (e.g., words, faces, foreign languages, skills), and have generally found evidence consistent with

---

<sup>2</sup>Ebbinghaus taught himself a list of 13 nonsense syllables (“nonsense” to rule out recall by association), and tested himself on the list at various time intervals. His metric for retention was a “saving rate,” – that is, the ratio of the number of repetitions needed to relearn the entire list and recite it twice from memory to the number of repetitions that were needed to learn and recite the list initially.

Ebbinghaus’s pioneering work. Levy and Loftus [48] remark:

“Many researchers, beginning with Ebbinghaus, have assumed forgetting to be exponential over time,  $t$ , between learning and test, that is:

$$P_t = e^{-kt}$$

Such a function would follow from the reasonable assumption that information in memory, like the content of many other physical systems, is lost at a rate that is proportional to the amount remaining in the system.”

Rubin and Wenzel [70], in a meta-study of 210 published data sets on retrospective recall, find that the exponential form  $p_t = be^{-mt}$  is one of the best 2-parameter fits of the retention function.<sup>3</sup>

### 1.2.2 Prospective Memory: Components and Types

Meacham and Singer [58] are credited with coining the term “prospective memory.” Cohen [18] describes the components of prospective memory as “remembering what the planned action is, remembering to perform it, and remembering when and where to do it.” It has been noted that PM tasks actually contain both a prospective and a retrospective component. Baddeley [7] defines the distinction between these components as “when” something should be remembered, versus “what.” For instance, if the task is to pass along a message to someone, the retrospective component is remembering the message, while the prospective component is remembering that we have a message for that person when we see them.

While PM tasks contain a retrospective component, they have a number of characteristics that distinguish them from pure retrospective recall. For one, PM tasks generally have low information content – Harris [41] notes that “the information to be recalled may be trivially easy, but remembering to recall at all may be the difficulty.” Also, while retrospective recall is cued or prompted by events in the present – e.g., being asked for someone’s phone number from memory, or partaking in an activity that requires the recall of

---

<sup>3</sup>Rubin and Wenzel find three other functional forms that fit well: logarithmic ( $p_t = b - m \log(t)$ ), power ( $p_t = bt^{-m}$ ), and hyperbolic ( $p_t = \frac{m}{t+b}$ )

some previously acquired skill – cues play much less of a role in prospective recall. While PM tasks are occasionally cued (e.g., driving by the grocery store and remembering that you have to buy groceries), these cues are generally random and thus affect the pattern of recall in a less predictable manner.<sup>4</sup>

### 1.2.3 Experimental Studies of Prospective Memory

Prior to Silk [77] and Ericson [27], most prospective memory experimental studies focused on the impact of exogenous factors on PM success rates (such as deadline length, subject age, and reminders provided by the experimenter), rather than testing for relevant PM belief and performance parameters in an economic decision-making framework. The most common were postcard studies, where subjects are asked to return postcards on certain dates or within some time interval (usually without incentives). Wilkins [88] asked 34 subjects to return one card each, from 2 to 36 days later, and found no effect of length of interval on performance. Meacham and Leiman [57], in a similar study, found that later cards were less likely to be posted than earlier cards, and that providing subjects with cues (such as colored tags on their key rings) improved return rates. Orne [64] and Meacham and Singer [58] found a similar effect of cues on return rates, and also found that monetary incentives improved return rates.<sup>5</sup> Since these early postcard studies, there has been a growing psychology literature testing subjects' prospective memory in "semi-naturalistic" (as opposed to laboratory) settings – see Table 7.2 in McDaniel and Einstein [56] for a summary of these experiments and the measured prospective memory success rates.

Silk [77] advanced this body of evidence by running a series of experiments that "examine consumers' purchase and post-purchase behavior with a real rebate offer." Silk lists a number of findings from the experiments, includ-

---

<sup>4</sup>Cues may affect the likelihood of recall, if not the pattern. That is, all other things equal, PM tasks with more frequent cues will be recalled more often. Additionally, one type of cue that can affect the pattern of recall is the use of memory aids (e.g., planners, Post-It notes, etc.), which we discuss in Section 1.6.

<sup>5</sup>Other studies (such as [76] and [34]) have focused on prospective memory with respect to appointment-keeping. Levy and Claravall [47], in a study of medical patients needing regular check-ups at varying intervals, found that reminders increased compliance most for patients with the longest intervals between appointments.

ing, 1) increasing the rebate reward increased takeup rates, but had “a weaker effect” on redemption rates; 2) increasing the length of the redemption period increased redemption confidence and takeup rates, but reduced redemption rates; and 3) there is a surprisingly positive correlation between randomized redemption effort levels and redemption rates (that is, redemption rates were higher in conditions where the rebate form was made more cumbersome and lengthy).

Another key area of research in the psychology literature has been the relationship between retrospective and prospective memory. In one of the best-known studies [89], medical subjects were tested on free recall of unrelated words, and then monitored in their pill-taking. Surprisingly, they find that subjects who did worse in the retrospective memory test (word recall) remembered to take their pills at a higher rate.<sup>6</sup> Other studies [92] [17], have found a positive correlation between performance in retrospective memory tests and PM tasks.

## 1.3 Baseline Model

### 1.3.1 Definitions

We consider an individual facing some task with cost  $c$  and benefit  $b$ , and  $T$  periods until an exogenously imposed deadline. In each period, there is some uncertainty about whether the individual will recall the task. An individual will execute the task in a period if 1) it is recalled; and 2) the expected utility from carrying out the task in that period is greater than the perceived expected utility from deferring the task and relying on future recall and execution.

The probability of recall is not fixed across periods – rather, it depends on the history of previous recall. Our model incorporates a simple form of the rehearsal property, where the probability of recall in some period  $t$  ( $p_t$ ) depends on the number of successive recall failures in the preceding periods.<sup>7</sup>

---

<sup>6</sup>The authors suggest this result is driven by the greater use of memory aids by the poor retrospective memory test performers, though they do not track the use of memory aids by the subjects.

<sup>7</sup>While the rehearsal property is more commonly associated with retrospective recall, it has

Define  $N_t$  as the number of successive, immediately preceding periods in which the task was not recalled (i.e., if the task was recalled at  $t - 1$ , then  $N_t = 0$ ; if the task was forgotten at  $t - 1$  but recalled at  $t - 2$ , then  $N_t = 1$ ; and so on). Then:

$$p_t = \theta^{N_t} p,$$

where  $p, \theta \leq 1$ . The term  $p$  captures the base rate of remembering – that is, the likelihood of remembering the task in the first period, and in any period preceded by a period of recall. The term  $\theta$  captures the factor by which memory decays (i.e., the factor by which the likelihood of recall is reduced) with each additional period of forgetting. Note that this form leads to exponential decay in expectation of future recall, consistent with the empirical evidence on retention discussed in Section 1.2. Our modeling of the rehearsal property is a slightly more nuanced version of that adopted in Mullainathan [60], which essentially assumes:

$$p_t = \begin{cases} p & \text{if } m_{t-1} = 1 \\ \theta p & \text{if } m_{t-1} = 0, \end{cases}$$

where  $m_t$  is a random variable equal to 1 if the task is recalled in period  $t$  and 0 otherwise (i.e.,  $p_t = E(m_t)$ ). While Mullainathan assumes that the recall probability is only contingent on recall in the immediately preceding period, we assume that each successive recall failure will lower the subsequent probability of recall by a factor  $\theta$ .

We use a search model structure, similar to O’Donoghue and Rabin [62], and assume the cost of task completion is stochastically determined each period. Let  $c$  be a random variable with CDF  $F(\cdot)$ , with support  $[\underline{c}, \bar{c}]$  and  $\underline{c} > 0$ . Let  $c_t$  denote the draw from this distribution at time  $t$  – that is, the cost of task execution at  $t$ . We assume the benefit is fixed at  $b$ , regardless of when the task is completed. In this section, we do not incorporate time-discounting into the agent’s preferences – whenever the agent completes the task, she receives utility  $b - c_t$ . An agent’s strategy is a sequence  $s = (s_1, s_2, \dots, s_T)$  where each  $s_t$  denotes the cost threshold for task execution. The task is executed in  $t$  if

---

a clear application to PM tasks. Having a task “top-of-mind” increases the chances that it will stay “top-of-mind” and be completed before the deadline. [24], in a study of a pulse PM task, found that “preservation” – the remembering that one has to carry out an action hours or even days prior to the planned time – improves the likelihood of correct recall.

the task is recalled in  $t$  and  $c_t \leq s_t$ . We denote first-best strategy for given memory parameters  $(p, \theta)$  by  $s^*(p, \theta) = (s_1^*(p, \theta), s_2^*(p, \theta), \dots, s_T^*(p, \theta))$ . This concept of a first-best strategy is referred to as “perception-perfect strategy” [62], which they define as “a strategy that in all periods (even those after the activity is performed) a person chooses the optimal action given her current preferences and her perceptions of her future behavior.”

Furthermore, define the following:

1.  $m_t$  as a random variable equal to 1 if the task is recalled in period  $t$  and 0 otherwise, with  $m = (m_1, \dots, m_T)$ . We will assume in this paper that  $m_1 = 1$ . Note that since  $m_t$  is binary the set of all possible recall outcomes is just  $2^T$ ;
2.  $\mu_t(m, p, \theta)$  as the probability from the time  $t$  perspective of memory sequence  $m$  conditional on  $m_t = 1$ ;
3.  $\pi_{t,t'}(s, p, \theta | m)$  as the probability from the time  $t$  perspective that the agent will not complete the task before  $t' > t$  given the memory sequence  $m$  if the agent chooses to defer the task in  $t$  and follows strategy  $s$  thereafter;
4.  $V_t(s, p, \theta)$  as the utility from the time  $t$  perspective of the task, given that it has not been completed before time  $t$ , and that the agent follows strategy  $s$ <sup>8</sup>.  $V_t$  is a random variable that depends on current and future values of  $c$  and  $m$ , and we will write its time  $t$  expectation in a Bellman equation below.

Note that

$$\pi_{t,t'}(s, p, \theta | m) = \begin{cases} 1 & \text{if } t' = t + 1 \\ \prod_{i=t+1}^{t'-1} 1 - m_i F(s_i) & \text{if } t' > t + 1. \end{cases}$$

That is, the probability of arriving in period  $t+1$  without having completed the task conditional on deferring in period  $t$  is 1, and for every following period

---

<sup>8</sup>We consider the effects of present-biased preferences on agent behavior in Section 1.6.

is the joint probability of drawing a cost above the threshold for each period that the task is recalled (i.e. where  $m_t = 1$ ).

We can write the following Bellman equation for  $E_t(V_t)$  by recognizing that the value of a task from the time  $t$  perspective is the maximum of exercising in  $t$  and deferring to  $t + 1$ , which sets up a recursive system.

$$\begin{aligned} V_t &= \max\{m_t(b - c_t), V_{t+1}\}, \text{ and} \\ E_t(V_t(s, p, \theta)) &= \max[m_t(b - c_t), E_t(V_{t+1}(s, p, \theta))]. \end{aligned}$$

The continuation value from deferring in period  $t$  is the second term in the max expression. It is equal to the probability of task execution multiplied by the expected utility conditional on task execution in each proceeding period from  $t + 1, \dots, T$ . Two alternative expressions for  $V_t$  involve  $V^1$  and  $V^0$  terms, one using same period  $V$ s and the other next period  $V$ s:

$$\begin{aligned} V_t(s, p, \theta) &= \max[m_t(b - c_t), pV_{t+1}^1 + (1 - p)V_{t+1}^0] \\ &= m_tV_t^1 + (1 - m_t)V_t^0 \end{aligned}$$

Finally, let  $Q(s, p, \theta)$  denote the ex ante probability of task execution over all  $T$  periods, such that

$$Q(s, p, \theta) = \sum_{t'=1}^T \sum_{m \in 2^T} \mu_t(m, p, \theta) \pi_{0,t'}(s, p, \theta | m) m_t F(s_t).$$

Define  $\hat{p}$  and  $\hat{\theta}$  as the agent's beliefs about the memory process. The agent will solve for her strategy  $s$  based on these beliefs by backwards induction. Note that in periods where  $m_t = 0$  and the agent forgets the task, the value of  $s_t$  is not relevant. So we define the optimal strategy only in periods for which  $m_t = 1$ :

$$s_t^*(\hat{p}, \hat{\theta}) = \begin{cases} b & \text{if } t = T, m_t = 1 \\ b - E_t(V_{t+1}(s, \hat{p}, \hat{\theta}) | m_t = 1) & \text{if } 1 \leq t < T, m_t = 1 \end{cases} \quad (1)$$

### 1.3.2 Some Results Involving Hazard Rates of Completion

We now consider how imperfect memory affects the time-pattern of hazard rates, where the hazard rate at  $t$  is defined as the probability of executing a



task at period  $t'$  without having executed the task. To do so, we first distinguish between agents with perfect memory, and agents with imperfect memory, but accurate beliefs about memory.

We define a “Perfect Rememberer” (PR) as an agent with  $p = \hat{p} = \theta = \hat{\theta} = 1$ , and agent with perfect memory and correct beliefs about memory.

We define a “Sophisticated Forgetter” (SF) as an agent with imperfect memory, that is with  $p < 1$  and  $\theta < 1$ , but correct beliefs about memory, that is  $p = \hat{p}$  and  $\theta = \hat{\theta}$ .

Define  $h_t$  to be the hazard rate of task completion at time  $t$ , and note that

$$h_t(s, p, \theta) = \text{Prob}(m_t = 1 \text{ and } c_t \leq s_t) = E(m_t)F(s_t)$$

That is, the hazard rate equals the probability of remembering the task at  $t$  conditional on forgetting the task or drawing a cost over the threshold in each previous period, multiplied by the probability that the current period cost is below the current period threshold.

**Theorem 1.** *For a PR, hazard rates strictly increase in time.*

This follows directly from the fact that  $s_t^*$ , the PR’s threshold cost, is increasing in time, which can be seen in the iterative solution for  $s^*$  in equation 1. Since  $E(m_t) = 1$  for all  $t$  for PRs, and  $F(s_t^*)$  is increasing in  $t$ , it follows that  $h_t(s^*, 1, 1)$  is increasing in  $t$ .

*Proof of Theorem 1.* We will show that  $s_t^*(1, 1)$ , the PR’s optimal  $s_t$ , is increasing in  $t$ . Since  $h_t = E(m_t)F(s_t)$  and  $F$  is increasing in  $s_t$ , this will establish the claim.

Note that a PR is defined by  $m = (1, 1, \dots, 1)$  with probability 1. We show that the continuation value to a PR of deferring in time  $t$ ,  $V_t$ , is decreasing in

$t$ :

$$\begin{aligned}
V_t &= \sum_{t'=t+1}^T \sum_{m \in 2^T} \mu_t(m) \pi_{t,t'}(\cdot | m) m_{t'} F(s_{t'}) [b - E(c_{t'} | c_{t'} \leq s_{t'})] \\
&= \sum_{t'=t+1}^T \pi_{t,t'}(\cdot | (1, \dots, 1)) F(s_{t'}) [b - E(c_{t'} | c_{t'} \leq s_{t'})] \\
&= \pi_{t,t+1}(\cdot | (1, \dots, 1)) F(s_{t+1}) [b - E(c_{t+1} | c_{t+1} \leq s_{t+1})] + V_{t+1},
\end{aligned}$$

Since the first term is always non-negative we see that  $V_t > V_{t+1}$  as claimed.

Then since  $s_t^* = b - V_t$ ,  $s_t^*$  is increasing in  $t$ .  $\square$

**Theorem 2.** *For any SF, with a sufficiently large  $T$  hazard rates will strictly decrease in time over some interval.*

The intuition for Theorem 2 is that as  $T$  grows large, the hazard in the last period converges to zero, because the fraction of the surviving population that remember in the last period goes to zero. However, the hazard in the first period is bounded below, even for arbitrarily large  $T$ .

*Proof of Theorem 2.* First, we argue that the hazard at a given number of periods away from the deadline converges to zero as the length of the deadline grows. In our notation,

$$\lim_{T \rightarrow \infty} h_{T-j} = 0, \forall j, p, \theta$$

The argument is as follows: consider an infinite population of identical SFs, and examine those that have not completed by the deadline. Note that the fraction of these individuals who complete in period  $T$  is by definition the hazard at time  $T$ . Some fraction will have  $m_T = 1$ , and those individuals complete at rate  $F(b)$ , since  $s_T^* = b$ . However, the rest with  $m_T = 0$  will complete at rate 0. Rerun this same experiment for a larger  $T$ . The fraction with  $m_t = 0$  will increase, and in fact will converge to 1 as  $T$  grows. This same argument holds for  $m_T - j$ , for each fixed  $j$ , as  $T$  grows. This argument establishes the claim.

We can prove this formally by noting that with strictly positive probability, an SF forgets and never remembers. We establish this in the following two

lemmas by noting that  $\{N_t\}$  forms a Markov Chain, and showing that 0 is a transient state:

**Lemma 1.** *The process returns to zero almost surely, i.e.  $P(N_{t+k} = 0 \text{ for some } k \geq 1 \mid N_t = 0) = 1$ , if and only if  $(1 - p_0)(1 - p_1)\dots(1 - p_n) \rightarrow 0$  as  $n \rightarrow \infty$ .*

*Proof.* Let  $R$  be the (random) time of the first return to zero, conditional on the process beginning at zero. Note that

$$P(N_{t+k} = 0 \text{ for some } k \geq 1 \mid N_t = 0) = 1$$

if and only if

$$P(R = \infty) = 0.$$

That is, the process returns to zero almost surely if and only if  $R$  is finite almost surely. Next, observe that

$$\begin{aligned} \{R > n\} \downarrow \{R = \infty\}, \text{ i.e.} \\ \{R > 1\} \supset \{R > 2\} \supset \dots \supset \bigcup_{n=1}^{\infty} \{R > n\} = \{R = \infty\}. \end{aligned}$$

So, by the downward continuity of the measure  $P$  we have

$$P(R > n) \downarrow P(R = \infty).$$

Thus

$$P(R = \infty) = 0$$

if and only if

$$P(R > n) \rightarrow 0.$$

The result follows from the fact that

$$P(R > n) = (1 - p_0)(1 - p_1)\dots(1 - p_{n-1}).$$

□

**Lemma 2.**  *$P(N_{t+k} = 0 \text{ for some } k \geq 1 \mid N_t = 0) < 1$ . That is, there is a strictly positive probability that the individual forgets forever.*

*Proof.* By Theorem 1, we need only show that

$$(1 - p_0)(1 - p_1)\dots(1 - p_n) = (1 - p)(1 - p\theta)\dots(1 - p\theta^n) \not\rightarrow 0$$

as  $n \rightarrow \infty$ .

This is equivalent to showing that

$$\log[(1 - p)(1 - p\theta)\dots(1 - p\theta^n)] \not\rightarrow -\infty \text{ as } n \rightarrow \infty,$$

which is in turn equivalent to showing that

$$\sum_{k=0}^{\infty} \log(1 - p\theta^k)$$

converges. Note that

$$\log(1 - p\theta^k) = -p\theta^k + O(\theta^{2k}) \text{ as } \theta^{2k} \rightarrow 0.$$

Thus

$$\sum_{k=0}^{\infty} \log(1 - p\theta^k) = \sum_{k=0}^{\infty} [-p\theta^k + O(\theta^{2k})],$$

and this series converges if and only if  $\theta < 1$ , which is true by assumption.  $\square$

Next, we claim that  $h_1$  for a SF is bounded below for all  $T$ . This will establish the result. For fix  $j$ , and suppose  $\underline{h}$  is the lower bound on  $h_1$ . By the above, for a sufficiently large  $T$  we can get  $h_{T-j} < \underline{h}$ . But then  $h_{T-j} < h_1$ , which shows the hazard must decrease at some point.

To show that  $h_1$  is bounded below, we show that  $s_1^*$  is bounded below, since  $h_1 = F(s_1^*)$ . The argument is as follows: suppose  $h_1$  were not bounded below. Then as  $T$  grows,  $h_1 \rightarrow 0$ . This implies that in the infinite-horizon problem the expected completion time is infinite; the intuition is that the SF waits around forever for an arbitrarily low cost draw. But this cannot be optimal, because in following this strategy the SF will at some point forget and never remember, with probability one. For we have argued above that 0 is a transient state of  $N_t$ , and so each period of remembrance carries a constant, positive probability that the SF will forget in the next period and never remember again. So the probability of such an eternal forgetting occurs with probability 1 in the infinite horizon problem. This implies that the expected utility from such a strategy is zero, which cannot be optimal. Therefore  $h_1$  must be bounded below.  $\square$

### 1.3.3 A Simple Example

To illustrate the solution employed by rational agents, and the resulting time pattern of hazard rates, we consider an example where  $T = 3, b = 1, c \sim U[0, 2]$ .

First consider the PR's backwards-induction reasoning. In the last period of the model, the PR will set  $s_3^* = 1$  and will thus execute the task in period 3, conditional on having arrived at period 3 without already completing the task, with probability 0.5.

Now consider the period 2 problem. The continuation value  $V_2(s^*)$  from deferring equals  $F(s_3^*)(b - E(c_3|c_3 \leq s_3^*)) = 0.5(1 - 0.5) = 0.25$ . Thus the PR will set  $s_2^* = 1 - 0.25 = 0.75$ .

Finally, in period 1 the PR will set  $s_1^* = b - F(s_2^*)(b - E(c_2|c_2 \leq s_2^*)) - (1 - F(s_3^*)F(s_3^*)(b - E(c_3|c_3 \leq b))) = 1 - (0.375(1 - 0.375)) - (0.625)(0.5)(1 - 0.5) = 0.61$ . This implies an ex ante expected utility  $V_0(s^*) = 0.48$  and the probability of task completion  $Q(s^*) = 0.78$ . The solution for the PR is summarized in Table 1.

Period	Optimal Threshold Cost ( $s_t^*$ )	Hazard Rate ( $h_t$ )	Contribution to Prob of Executing	Contribution to Task's Expected Utility
1	0.61	0.31	0.31	0.21
2	0.75	0.38	0.26	0.16
3	1.00	0.50	0.21	0.11
Total			0.78	0.48

Table 1: PR Behavior for  $b = 1, c_t \sim U[0, 2], T = 3$

Now consider a SF. Assume the memory process is characterized by  $p = 0.5, \theta = 0.2$ . In period 3, conditional on recall the SF acts just like the PR: as long as the realization of  $c_3$  is less than or equal to  $b$ , she will execute that task. In our notation she uses  $s_3^*(0.5, 0.2) = s_3^*(1, 1) = b = 1$ .

However, in period 2 the SF realizes that her continuation value is lower due to imperfect memory and uses  $s_2^*(0.5, 0.2) > s_2^*(1, 1)$  accordingly. In particular  $V_2(s^*(p, \theta), p, \theta) = pV_2(s^*(1, 1), 1, 1) = (0.5)(0.25) = 0.125$ , and so

$$s_2^* = 0.875.$$

In period 1, the SF uses  $s_1 = b - V_1(s)$ , where

$$\begin{aligned} V_1(s) &= \sum_{t'=2}^3 \sum_{m \in 2^3} \mu_1(m, p, \theta) \pi_{1,t'}(s, p, \theta | m) m_t F(s_1) (b - E(c_{t'} | c_{t'} \leq s_{t'})) \\ &= pF(s_2)(b - E(c|c \leq s_2)) + [(1-p)\theta p + p^2(1-F(s_2))]F(s_3)(b - E(c|c \leq s_3)) \\ &= (0.5)(0.4375)(1 - 0.4375) + (0.05 + 0.14)(0.5)(1 - 0.5) \\ &= 0.17. \end{aligned}$$

So,  $s_1^* = 0.83$ . this implies an ex ante expected utility  $V_0(s) = 0.19$  and a probability of task completion  $Q(s) = 0.33$ . Table 2 summarizes SF behavior.

Period	Optimal Threshold Cost ( $s_t^*$ )	Hazard Rate ( $h_t$ )	Contribution to Prob of Executing	Contribution to Task's Expected Utility
1	0.83	0.21	0.21	0.12
2	0.87	0.11	0.08	0.05
3	1.00	0.06	0.04	0.02
Total			0.33	0.19

Table 2: SF Behavior for  $b = 1, c_t \sim U[0, 2], T = 3, p = 0.5, \theta = 0.2$

Note that in this particular example, hazard rates are decreasing in time for the SF and increasing for the PR.

Figure 1 shows the sensitivity of the ex ante probability of task completion  $Q$  to  $p$  and  $\theta$ . Figure 2 shows the sensitivity of  $V_0$  to  $p$  and  $\theta$ .

## 1.4 Overconfidence in Prospective Memory

Thus far we have assumed that agents, despite their fallible memory, have correct beliefs over their prospective memory processes, and account for this perfectly in calculating threshold costs each period. We now discuss augmenting the model by allowing for systematic errors in prospective memory beliefs. Specifically, we consider prospective memory overconfidence – in the context of the model,  $\hat{p} > p$  and/or  $\hat{\theta} > \theta$ .

Figure 1: Sensitivity of  $Q$  to  $p$  and  $\theta$ .

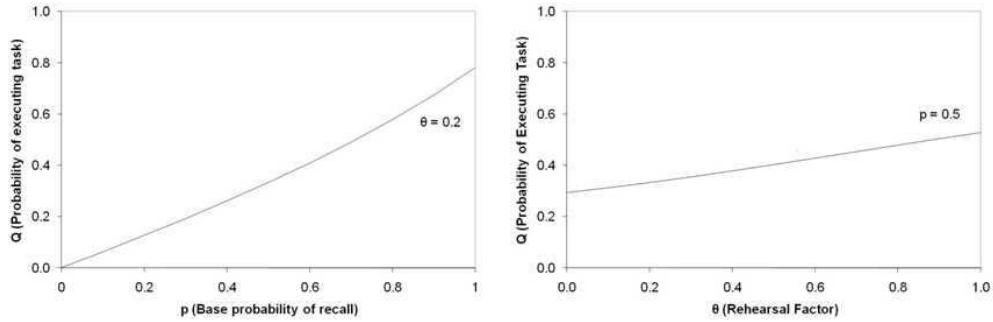
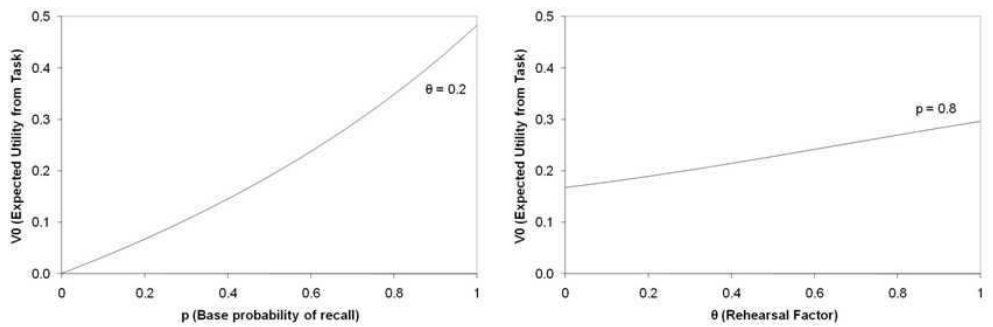


Figure 2: Sensitivity of  $V_0$  to  $p$  and  $\theta$ .



#### 1.4.1 Overconfidence Literature

Experimental and survey studies have found that the average individual exhibits overconfidence over a wide range of skills, abilities, and personal traits. Most people believe they are more intelligent [93], more sociable and better leaders [19], more productive [66], and better drivers than the average person [82]. To the extent that remembering things is a skill or ability like those above, assuming overconfidence in memory is a natural extension of these findings.

Most of the evidence on memory overconfidence, until recently, has been in retrospective contexts – that is, individuals are overconfident in the accuracy

of their recollections.<sup>9</sup> For instance, eyewitnesses tend to be overconfident in the accuracy of lineup identifications of criminal suspects, and the details of their trial testimony (see [52] and [10]). However, two recent studies – Silk [77] and Ericson [27] – find evidence of prospective memory overconfidence in experimental designs that broadly fit the task model presented in Section 1.3.

Silk contains an experimental condition where subjects are offered a rebate and are asked for their “redemption confidence” (i.e., a subjective probability of redemption), as well as their expectation of the overall redemption rate in the population (i.e., a base rate estimate). Those who choose to accept the rebate offer are, on average, 93.5% confident they will submit it, while the ultimate redemption rate is only 60%. Their population (or base rate) estimates are much more in line with the true redemption rates, suggesting that prospective memory is yet another trait where individuals systematically evaluate themselves as being “above average.” On the other hand, those who reject the rebate offer have subjective probability rates that are slightly lower than both their base rate estimates and the eventual redemption rate, suggesting heterogeneity in prospective memory overconfidence.

Ericson uses an incentive-compatible design to elicit subjects’ beliefs that they will remember to claim a future contingent payment, then compares these beliefs with actual claim behavior. Subjects are asked for their preference between a conditional payment of \$20 (contingent on the subject sending an email to the experimenter in a 5-day window 3-4 months in the future) and an automatic payment \$x, where x varies from \$5 to \$20 in \$0.75 increments (each subject gives their preference between the contingent payment and the automatic payment for each potential value of x). Subjects are then randomly assigned to one of three conditions – 1) receiving a \$20 automatic payment; 2) receiving a \$20 contingent payment; or 3) the choice they made between the \$20 contingent payment and a \$x automatic payment for a randomly selected value of x. The threshold values of \$x for which subjects switch their preference from the contingent payment to the automatic payment are used to infer subjects’ beliefs about their likelihood of claiming the contingent payment. The claim probability implied by the preferences of the subjects in the contingent payment condition was 0.76, while the actual claim rate was 0.53 –

---

<sup>9</sup>Mullainathan [60] makes a related assumption that individuals have limited memories, but assume their recollections to be perfectly accurate.



a significant difference at the 99.9% level and evidence of prospective memory overconfidence in the subject pool.

#### 1.4.2 Overestimating $p$ vs Overestimating $\theta$

Overestimating  $p$  can be interpreted as being overconfident in the base level of future recall probability. In the extreme case, when  $\hat{p} = 1$ , the agent believes that she will recall the task with probability 1 in every future period – that is, that she believes she is a perfect rememberer. Cohen [18] notes that “people’s beliefs about their own [prospective] memories are based on their experience and success and failure in everyday life.” In this light, overestimating  $p$  can be viewed as a “recency bias” – an agent who correctly remembers the task in period  $t$  may upwardly revise PM beliefs above the true value of  $p$  for periods  $t + 1, \dots, T$ .

Overestimating  $\theta$  can be interpreted as underestimating the rate at which a specific memory weakens once forgotten. In the extreme case, when  $\hat{\theta} = 1$ , the agent believes the recall probability in all future periods is identical (and equal to  $\hat{p}$ ). Overestimating  $\theta$  can be interpreted as a form of projection bias, as defined by Loewenstein, O’Donoghue, & Rabin [51]. While these authors use the term to describe the tendency to project current preferences on future selves, in this case agents project the current memory state onto future selves. When an agent is in a state of recall, as she must be whenever assessing whether to perform the task or not, the likelihood of recall in the next period is  $p$ . If an agent believes  $\hat{\theta} > \theta$ , she underestimates the effect that forgetting will have on eroding her current memory state – that is, she estimates the probability of recall for every possible  $N_t$  to be closer to the next period probability of recall than it actually is. As with agents exhibiting projection bias, agents overestimating  $\theta$  underappreciate how a change in circumstances will affect some parameter.

Despite these different interpretations, the effects of overestimating either parameter in our model are similar – in both cases, agents overestimate the continuation value of deferral and set inefficiently low threshold costs – and thus we do not emphasize the distinction in what follows.

### 1.4.3 PM Overconfidence as Information Projection

Both instances of prospective memory overconfidence – that is, overestimating  $p$  or  $\theta$  – are closely related to other examples of “information projection” such as hindsight bias<sup>10</sup> and curse of knowledge<sup>11</sup>. Hindsight bias is the tendency to feel like we knew the outcome of some probabilistic event (e.g., a football game, political election, or business investment) all along. It is generated by the inability to fathom ever being without the information that we have after the fact – i.e., the inability to imagine having been in a different information state. PM overconfidence is similarly caused by the difficulty of imagining being without information we have in the present, only now projecting that information state on the future rather than the past.

“Curse of Knowledge” refers to the inability to ignore information in making economic decisions, even when internalizing that information is harmful. Thaler [83] writes of the curse of knowledge, “once we know something, we can’t ever imagine thinking otherwise. This makes it hard for us to realize that what we know may be less than obvious to others who are less informed.” In the case of prospective memory overconfidence, the (potentially) less informed “others” are our future selves. We find it difficult to imagine not knowing (or remembering) what we know (or are aware of) in the present, and thus are overoptimistic in our capacity to later recall information that is currently top of mind.

## 1.5 Effects of PM Overconfidence on Agent Behavior

We can immediately establish several results about PM overconfidence and ex ante task completion and welfare:

**Theorem 3.** *For any agent with imperfect memory and overconfident beliefs, the ex ante likelihood of task completion decreases in the level of overconfidence.*

*Proof of Theorem 3.* The statement is that  $Q(s^*, \hat{p}, \hat{\theta})$  decreases in  $\hat{p}$  and  $\hat{\theta}$  for  $\hat{p} \leq p, \hat{\theta} \leq \theta$ .

---

<sup>10</sup>See [29]

<sup>11</sup>See [11]

First we will show that  $s_t(p, \theta)$  is decreasing in  $p$  and  $\theta$ , by induction on  $t$ . Recall that  $s_T = b$ , and observe that  $s_{T-1} = b - E_{T-1}(V_T | m_{T-1} = 1)$ . Then note that

$$\begin{aligned} E_{T-1}(V_T | m_{T-1} = 1) &= E_{T-1}(\max(m_T(b - c_T), V_{T+1} | m_{T-1} = 1)) \\ &= E_{T-1}(m_T(b - c_T) | m_{T-1} = 1) \\ &= p(b - E_{T-1}(c_T)) \end{aligned}$$

which is clearly increasing in  $p$ . Next suppose that  $s_t$  is decreasing in  $p, \theta$ . This implies that  $E_t(V_{t+1} | m_t = 1)$  is increasing in  $p, \theta$ . Recall that  $s_{t-1} = b - E_{t-1}(V_t | m_{t-1} = 1)$  and  $V_t = \max(m_t(b - c_t), V_{t+1})$ , so

$$\begin{aligned} E_{t-1}(v_t | m_{t-1} = 1) &= E_{t-1}(\max(m_t(b - c_t), V_{t+1} | m_{t-1} = 1)) \\ &= \max(p(b - E_t - 1(c_t)), E_{t-1}(V_{t+1} | m_{t-1} = 1)) \\ &= \max(p(b - E_t - 1(c_t)), pE_t(V_{t+1} | m_t = 1, m_{t-1} = 1) + \\ &\quad (1 - p)E_t(V_{t+1} | m_t = 0, m_{t-1} = 1)) \end{aligned}$$

The first term in the max is clearly increasing in  $p$ . Consider the second term,  $pE_t(V_{t+1} | m_t = 1, m_{t-1} = 1) + (1 - p)E_t(V_{t+1} | m_t = 0, m_{t-1} = 1)$ . The first term is increasing in  $p, \theta$  by the induction hypothesis. The second term is as well, since we can continue writing  $V_{t+1}$  in terms of  $V_{t+2}$  and so forth, using the induction hypothesis each time to deal with terms of the form  $E_{t-j}(V_{t-j+1} | m_{t-j} = 1)$ . The only term we will need to deal with is of the form  $E_{T-1}(V_T | m_{T-1} = 0, m_{T-2} = 0, \dots, m_{t-1} = 1)$ , which is clearly increasing in  $p, \theta$ . Finally, note that  $E_t(V_{t+1} | m_t = 1, m_{t-1} = 1) > E_t(V_{t+1} | m_t = 0, m_{t-1} = 1)$ , so increasing its weight, namely  $p$ , will increase the entire quantity. The same argument applies for the weights that expand out of the second term

Therefore  $s_t(p, \theta)$  is decreasing in  $p$  and  $\theta$ . Since an overconfident agent follows the exact same strategy as a sophisticated forgetter with the same  $\hat{p}, \hat{\theta}$ , and since we have just established that her reservation costs will be uniformly lower than an agent with the same true memory parameters but less overconfidence, we have established that the probability of task completion,  $Q(s^*, \hat{p}, \hat{\theta})$ , is decreasing in  $\hat{p}$  and  $\hat{\theta}$ .  $\square$

**Theorem 4.** *For any agent with imperfect memory and overconfident beliefs, the ex ante expected utility of a task is decreasing in the level of overconfidence. Here we assume the ex ante utility is computed knowing the true memory parameters, and independently of the memory beliefs.*

The intuition for these results is that agents with PM overconfidence overestimate the continuation value of task deferral, and thus set cost thresholds below the optimal level every period before  $T$ . In our notation,  $s_t(\hat{p}, \hat{\theta}) < s_t(p, \theta) \forall t < T$ . The lower probability of task completion follows immediately from following a strategy with lower threshold costs (holding the true memory parameters constant).

To see that utilities are lower, note that these lower threshold costs imply that for every period before  $T$ , an overconfident agent defers the task for a range of cost realizations when the first-best strategy would dictate executing the task. Thus, it must be that  $V_0(s_t(\hat{p}, \hat{\theta}), p, \theta) \leq V_0(s(p, \theta), p, \theta)$ . More generally, cost thresholds are decreasing in the level of PM overconfidence, as we show in the proof.

*Proof of Theorem 4.* The statement is that  $E_0(V_0(s^*, \hat{p}, \hat{\theta}))$  decreases in  $\hat{p}$  and  $\hat{\theta}$  for  $\hat{p} \leq p, \hat{\theta} \leq \theta$ .

From the previous proof, we have shown that as  $(\hat{p}, \hat{\theta})$  go to  $(1, 1)$ , the strategy of the NF converges monotonically away from that of the PR and towards that of the PR. We claim (without proof) that as the NF's strategy moves monotonically away from the PR's strategy as  $(\hat{p}, \hat{\theta})$  go to  $(1, 1)$ , that the utility loss increases monotonically as well. Put another way, the farther the strategy is from the optimal, the greater will be the utility loss, and we have established that the distance from the NF's strategy to the optimal PR's strategy increases in the degree of naivete.  $\square$

In comparing the behavior of agents with PM biases to rational agents (PRs and SFs), we find it useful to define an agent exhibiting extreme bias with respect to PM beliefs:

We define a “Naive Forgetter” as an agent with  $p < 1, \theta < 1$  but  $\hat{p} = 1$ . NFs incorrectly anticipate perfect recall in all future periods. Note that for NFs,  $s = s^*$  – that is, they employ the same set of threshold costs as PRs, only this strategy is inefficient given their true memory parameters.

### 1.5.1 Some Results Involving Length of Deadline

We now investigate the effect of lengthening the deadline on task completion for the PR, SF, and NF. We update the notation by adding time subscripts to  $s, s_t, h, V_t$ , and  $Q$  to represent the number of periods in the particular problem.

**Theorem 5.** *For a PR, the probability of task completion and the expected utility of the task are strictly increasing in the time allotted. In our subscripted notation, we claim that  $V_{0,T+\delta}(s_{T+\delta}^*) > V_{0,T}(s_T^*)$  and  $Q_{T+\delta}(s_{T+\delta}^*) > Q_T(s_T^*)$  for all  $T, \delta$ .*

**Theorem 6.** *For a SF, the expected utility of the task is strictly increasing in the time allotted for any  $p > 0$ .*

**Theorem 7.** *For a NF, given any two deadline lengths  $T$  and  $T + \delta$ ,*

1. *There exist sufficiently poor memory parameters such that the probability of task completion is lower for the longer deadline. That is, there exists some  $p', \theta'$  such that for  $0 < p < p'$  and  $\theta < \theta'$ ,  $Q_{T+\delta}(s_{T+\delta}^*, p, \theta) < Q_T(s_T^*, p, \theta)$*
2. *There exist sufficiently poor memory parameters such that the ex ante expected utility from task completion is lower for the longer length of deadline. That is, there exists some  $p', \theta'$  such that for  $0 < p < p'$  and  $\theta < \theta'$ ,  $V_{0,T+\delta}(s_{T+\delta}^*, p, \theta) < V_{0,T}(s_T^*, p, \theta)$ .*

For the intuition behind this result, consider the case where  $p$  is small and  $\theta = 0$ . In this case, the overall probability of task execution is approximately equal to the probability of task execution in the first period, since the probability of task execution at  $t = 1$  is proportional to  $p$ , and for all other periods, is proportional to some higher order power of  $p$ . In particular, the ex ante probability of completing the task in period  $k$  is proportional to  $p^k$ . But the probability of task execution in period 1,  $pF(s_{1,T}^*)$ , is decreasing in the length of deadline  $T$ , since  $s_{1,t+\delta}^* < s_{1,t}^*$  implies  $F(s_{1,t+\delta}^*) < F(s_{1,t}^*)$ . Thus, for  $p$  small and  $\theta = 0$ ,  $Q_{T+\delta}(s_{T+\delta}^*, p, \theta) \approx pF(s_{1,T+\delta}^*) < pF(s_{1,T}^*) \approx Q_T(s_T^*, p, \theta)$ . Since  $Q_{T+\delta}(s_{T+\delta}^*, p, \theta)$  and  $Q_T(s_T^*, p, \theta)$  are continuous functions of  $p$  and  $\theta$ ,  $Q_{T+\delta}(s_{T+\delta}^*, p, \theta) - Q_T(s_T^*, p, \theta)$  is continuous in  $p$  and  $\theta$  and thus the result holds for some range of sufficiently  $p, \theta$ .

The analogous result for ex ante expected utility can be demonstrated in a similar manner. The ex ante expected utility from executing in period  $1 - pF(s_{1,T}^*)E(c|c < s_{1,T}^*) -$  approximates overall expected utility for small  $p$  and  $\theta = 0$ , since it is proportional to  $p$  and expected utility contributions for all future periods are proportional to some higher order power of  $p$ . As long as  $s_{1,t}^* < b$  (which must always hold), the ex ante expected utility from executing in period 1 is decreasing in  $t$ , since  $s_{1,t}^*$  is decreasing in  $t$ . Again, by the continuity of  $V_{0,T+\delta}$  and  $V_{0,T}$  in  $p, \theta$ , their difference is continuous in  $p, \theta$ , thus expected utility for a NF is decreasing in the length of deadline for some range of sufficiently small  $p, \theta$ .

*Proof of Theorem 5.* This theorem becomes obvious once one observes that the structure of the last  $T$  periods of a  $T + \delta$ -deadline problem is identical to the structure of a  $T$ -deadline problem. That is,  $s_{T+\delta, t+\delta}^* = s_{T,t}^*$  for  $t \in \{1, 2, \dots, T\}$ . Thus the probability of completing in periods  $1 + \delta, 2 + \delta, \dots, T + \delta$  with a  $T + \delta$ -deadline, conditional on having not completed until  $1 + \delta$ , is the same as the probability of completing in the  $T$ -deadline case. Since the probability of completing in periods  $1, 2, \dots, \delta$  is non-negative, it is clear that  $Q_{T+\delta}(s_{T+\delta}^*) > Q_T(s_T^*)$ .

As for the utility from the task, a similar argument holds – since the optimal strategy in  $1 + \delta, 2 + \delta, \dots, T + \delta$  for the  $T + \delta$  problem is the same as in the original  $1, 2, \dots, T$  problem, the utility from those periods conditional on arriving is the same as the total utility of the original problem. Then setting  $s_1, s_2, \dots, s_\delta$  equal to  $b$  in the longer deadline case will result in the same utility for the two problems, and by letting them be arbitrarily close but strictly below  $b$  we can add on strictly positive utility.  $\square$

*Proof of Theorem 6.* The argument is similar to the utility portion of Theorem 5. However, while we before relied on the fact that the PR's optimal strategy is identical in the ending periods of an extended deadline strategy, we will now only rely on the fact that an SF *could* use the same strategy in an extended deadline setting.

Specifically, consider extending from a  $T$  deadline to a  $T + \delta$  deadline. An SF could use  $s_T^*$  for the first  $T$  periods of her  $s_{T+\delta}$  strategy. This would result in an ex ante utility obtained during those periods equal to the total ex ante

utility from a  $T$  deadline problem. The addition of additional ex ante utility in the final  $\delta$  periods would give the utility nod to the  $T + \delta$  problem. The fact that the SF will not in general actually use this replicated strategy does matter; all we argue is that the replicated strategy presents a lower bound on utility.  $\square$

*Proof of Theorem 7.* We begin with notation. Let  $Q_{T,t}$  be the ex ante probability that the first task completion occurs in period  $t$  of a  $T$ -deadline problem. Then

$$Q_T = \sum_{t=1}^T Q_{T,t}. \text{ Note that}$$

$$\begin{aligned} Q_{T,1} &= P(m_1 = 1)P(c_1 < s_{T,1}) \\ &= pF(s_{T,1}^*), \end{aligned}$$

$$\begin{aligned} Q_{T,2} &= P(m_2 = 1)P(c_1 < s_{T,1})(1 - Q_{T,1}) \\ &= [P(m_2 = 1|N_2 = 0)P(N_2 = 0) + P(m_2 = 1|N_2 = 1)P(N_2 = 1)]F(s_{T,2}^*)(1 - Q_{T,1}), \\ &= \left[ p \frac{1 - F(s_{T,1})}{1 - pF(s_{T,1})} p + p\theta \frac{1}{1 - pF(s_{T,1})} (1 - p) \right] F(s_{T,2}^*)(1 - Q_{T,1}), \text{ and in general} \end{aligned}$$

$$\begin{aligned} Q_{T,t} &= P(m_t = 1)P(c_t < s_{T,t}^*) \prod_{j=1}^{t-1} (1 - Q_{T,j}) \\ &= \left[ \sum_{k=0}^{t-1} P(m_t = 1|N_t = k)P(N_t = k) \right] F(s_{T,t}^*) \prod_{j=1}^{t-1} (1 - Q_{T,j}) \end{aligned}$$

Recall that  $P(m_t = 1|N_t = k) = \theta^k p$ . Also note that  $P(N_t = 0)$  is of the form  $\sum_{i=0}^{t-1} \sum_{j=0}^{t-2} a_{ij} p^i \theta^j$  where  $a_{i0} > 0$  if and only if  $i = t - 1$ . In fact we have  $a_{t-1,0} = 1$ , since the only term of the double sum contributing to  $N_t = 0$  involving  $p^{t-1}$

and  $\theta^0$  corresponds to the case where  $m_1 = 1, m_2 = 1, \dots, m_{t-1} = 1$ . Thus

$$\begin{aligned}
\lim_{\theta \rightarrow 0} Q_{T,t} &= \lim_{\theta \rightarrow 0} \left[ \sum_{k=0}^{t-1} \theta^k p P(N_t = k) \right] F(s_{T,t}^*) \prod_{j=1}^{t-1} (1 - Q_{T,j}) \\
&= p F(s_{T,t}^*) \lim_{\theta \rightarrow 0} \sum_{i=0}^{t-1} \sum_{j=0}^{t-2} a_{ij} p^i \theta^j \prod_{j=1}^{t-1} (1 - Q_{T,j}) \\
&= p F(s_{T,t}^*) \lim_{\theta \rightarrow 0} \sum_{i=0}^{t-1} a_{i0} p^i \prod_{j=1}^{t-1} (1 - Q_{T,j}) \\
&= p^t F(s_{T,t}^*) \lim_{\theta \rightarrow 0} \prod_{j=1}^{t-1} (1 - Q_{T,j})
\end{aligned}$$

From these recursions we see that  $\lim_{\theta \rightarrow 0} Q_{T,t}$  is in fact a polynomial in  $p$  and that the order of this polynomial is strictly greater than 1 except in the case where  $t = 1$ , and then the order is exactly 1. Add these per-period completion rates to form the total completion rate as

$$\begin{aligned}
\lim_{\theta \rightarrow 0} Q_T &= \lim_{\theta \rightarrow 0} \sum_{t=1}^T Q_{T,t} \\
&= \sum_{t=1}^T \lim_{\theta \rightarrow 0} Q_{T,t} \\
&= \sum_{t=1}^T p^t F(s_{T,t}^*) \lim_{\theta \rightarrow 0} \prod_{j=1}^{t-1} (1 - Q_{T,j})
\end{aligned}$$

and then compare  $Q_T$  to  $Q_{T+\delta}$  in the limit as  $\theta \rightarrow 0$ :

$$\lim_{\theta \rightarrow 0} \frac{Q_T}{Q_{T+\delta}} = \frac{\sum_{t=1}^T p^t F(s_{T,t}^*) \lim_{\theta \rightarrow 0} \prod_{j=1}^{t-1} (1 - Q_{T,j})}{\sum_{t=1}^{T+\delta} p^t F(s_{T+\delta,t}^*) \lim_{\theta \rightarrow 0} \prod_{j=1}^{t-1} (1 - Q_{T+\delta,j})}$$

Next we take the limit of both sides as  $p \rightarrow 0$ , use the fact discussed above that  $p^t F(s_{T,t}^*) \lim_{\theta \rightarrow 0} \prod_{j=1}^{t-1} (1 - Q_{T,j})$  is a polynomial in  $p$ , and apply L'Hospital's rule



to obtain

$$\lim_{p, \theta \rightarrow 0} \frac{Q_T}{Q_{T+\delta}} = \frac{F(s_{T,1}^*)}{F(s_{T+\delta,1}^*)}$$

In fact  $s_{T,1}^* = s_{T+\delta,1+\delta}^* > s_{T+\delta,1}^*$ ; see the proofs for Theorems 5 and 1 for the equality and inequality, respectively. Therefore

$$\lim_{p, \theta \rightarrow 0} \frac{Q_T}{Q_{T+\delta}} > 1.$$

Then since  $Q_T$  is continuous in  $p$  and  $\theta$ , we can find  $p', \theta'$  such that  $0 < p < p', 0 < \theta < \theta'$  implies that  $\frac{Q_T(s_T^*, p, \theta)}{Q_{T+\delta}(s_{T+\delta, p, \theta}^*)} > 1$ , i.e. that  $Q_T(s_T^*, p, \theta) > Q_{T+\delta}(s_{T+\delta, p, \theta}^*)$ . This establishes Part 1. of the Theorem.

For Part 2., we can use the intermediate result of Part 1. that as  $p, \theta \rightarrow 0$  the fraction of task completion occurring during the first period converges to 1. Combine this with the fact that  $s_{T,1}^* > s_{T+\delta,1}^*$  and we see that in the limit, the set of costs for which the agent in the  $T$ -deadline problem completes the task strictly contains those for which the  $T + \delta$ -deadline agent completes, and that the additional costs for which she completes add strictly positive utility. This establishes Part 2.  $\square$

### 1.5.2 Two- to Three- Period Example

Table 3 shows the change in the probability of task completion and expected utility for SFs as the number of periods is increased from 2 to 3 (for  $p = 0.5, \theta = 0.2, b = 1, c \sim U[0, 2]$ ). Consistent with Theorem 6, expected utility increases for all combinations of  $p, \theta$ .

Table 4 shows the change in the probability of task completion and expected utility for NFs as the number of periods is increased from 2 to 3. Consistent with Theorem 7, the probability of task completion and expected utility decrease for low values of  $p, \theta$ .

p/ $\theta$	Change in Probability of Task Completion					Change in Expected Utility				
	0.0	0.2	0.5	0.8	1.0	0.0	0.2	0.5	0.8	1.0
0.2	0.000	0.005	0.023	0.051	0.074	0.001	0.004	0.014	0.029	0.041
0.5	0.005	0.019	0.051	0.090	0.114	0.009	0.019	0.039	0.062	0.076
0.8	0.037	0.051	0.077	0.101	0.110	0.044	0.053	0.069	0.084	0.090
1.0	0.096	0.096	0.096	0.096	0.096	0.093	0.093	0.093	0.093	0.093

Table 3: Change in Probability of Task Completion and Expected Utility for SF, 3 Periods vs 2 Periods,  $b = 1, c_t U[0, 2]$

p/ $\theta$	Change in Probability of Task Completion					Change in Expected Utility				
	0.0	0.2	0.5	0.8	1.0	0.0	0.2	0.5	0.8	1.0
0.2	(0.014)	(0.011)	0.004	0.030	0.051	(0.003)	0.000	0.010	0.025	0.037
0.5	(0.021)	(0.006)	0.028	0.068	0.093	0.004	0.014	0.034	0.058	0.073
0.8	0.022	0.037	0.066	0.092	0.102	0.042	0.051	0.068	0.083	0.090
1.0	0.096	0.096	0.096	0.096	0.096	0.093	0.093	0.093	0.093	0.093

Table 4: Change in Probability of Task Completion and Expected Utility for NF, 3 Periods vs 2 Periods,  $b = 1, c_t U[0, 2]$

## 1.6 Extensions of the Model

### 1.6.1 Memory Aids and Reminders

Sophisticated agents accurately recognize the welfare loss from their fallible memory. As such, they may have a demand for memory aids which, at some cost, ensure (or increase the likelihood of) remembering in future periods. This can be thought of as analogous to sophisticated hyperbolic discounters' demand for commitment devices.

We define a memory aid  $a = (a_1, a_2, \dots, a_T)$ , where  $a_t \in \{0, 1\}$  for all  $t \in \{1 \dots T\}$ . If  $a_t = 1$  then the agent remembers the task in period  $t$  with probability 1. Define  $a^1$  as the complete memory aid such that  $a_t^1 = 1$  for all  $t \in \{1 \dots T\}$ , and  $a^0$  as the “no aid” choice such that  $a_t^0 = 0$  for all  $t \in \{1, \dots, T\}$ . Let  $A$  denote the set of all  $2^T$  possible memory aids, and  $\kappa(a)$  denote the utility-price of memory aid  $a$ , with  $\kappa(a^0) = 0$ .

We assume that in period 0 the agent is presented with the full menu of memory aids  $A$  and prices  $\kappa(a)$  for each  $a \in A$ . The agent can choose to implement any memory aid at cost  $\kappa(a)$ , and is then completely committed to

that memory aid for the full length of the problem. A memory aid cannot be adjusted at any time after  $t = 0$ .

We now define the following memory-aid-augmented versions of  $V_t$  and  $s$ :

1.  $U_t(s, p, \theta, a)$  as the continuation value at  $t$  from executing strategy  $s$  given memory parameters  $p, s$ , and memory aid  $a$ ; and
2.  $\psi(p, \theta, a)$  as the first-best strategy given  $p, \theta, a$ .

Note that  $\psi(p, \theta, a^0) = s(p, \theta)$  and  $U_t(\psi(p, \theta, a^0), p, \theta, a^0) = V_t(s(p, \theta), p, \theta)$ .

Given a complete menu of memory aids at time 0, define  $a^*$  as the memory aid chosen by the SF. Since the SF correctly perceives her lack of memory, we can define this memory aid as one satisfying the following property:

$$a^* = \arg \max_{a \in A} [U_0(\psi(p, \theta, a), p, \theta, a) - \kappa(a)]$$

The following Theorem examines the welfare loss to a SF of her forgetting.

**Theorem 8.** *With a complete menu of memory aids, a SF's welfare loss relative to that of an otherwise identical PR is bounded by  $\kappa(a^1)$  as  $b \rightarrow \infty$ .*

When one views  $\kappa(a^1)$  as the cost of upgrading to perfect memory, this theorem is obvious.

*Proof of Theorem 8.* Trivial. For a utility-price of  $\kappa(a^1)$  the PR can turn herself into a SF. Thus utility loss of the PR relative to the SF is bounded by this amount.  $\square$

Memory aids provide three benefits for a forgetful agent. First, they increase the ex ante expected utility contribution for every period with a reminder, since they increase the probability of recall in those periods to 1. But they provide two other subtler, yet important benefits. Memory aids allow agents to “gamble” with lower threshold costs in any period before a reminder. Consider the case where  $p$  is small. Without memory aids, threshold costs will

be close to  $b$  in every period, since the SF recognizes the small chance of recall. However, with any reminder, the SF knows that she will remember the task for sure in some period, and thus can more aggressively set a threshold cost away from  $b$  without worrying about wasting an opportunity to execute the task. The third benefit of a memory aid is that it resets the memory process to the state of recall (i.e., a reminder at  $t$  implies  $N_{t+1} = 0$ ), and thus increases the chances of recall in all future periods. This benefit could be thought of as keeping a low  $\theta$  from “kicking in” – that is, mitigating the adverse effect of repeated forgetting on the probability of recall.

What does the optimal memory aid look like, in terms of the number and sequencing of reminders? It depends on the relative size of these three effects. The first two effects suggest that if the price of reminders across time is constant, later reminders will be preferred to earlier reminders. Setting the probability of recall in any period to 1 is more beneficial in later periods, since it buys more recall. By this we mean that without memory aids, the ex ante probability of recall is declining in time, and so a later reminder increases that period’s recall by a larger amount than an earlier reminder would. Furthermore, later reminders provide agents with more periods to gamble with low threshold costs.

However, the third effect, that is, mitigating the effect of  $\theta$  on the likelihood of recall, points to another reminder structure. In particular, this effect would induce an agent to implement an intermittent reminder, where the agent spreads reminders out as much as possible, reducing the welfare loss caused by successive forgetting. For instance, if an agent can choose any  $T/2$  reminders, choosing reminders every other period will eliminate  $\theta$  from the problem entirely. Thus, for low values of  $\theta$ , this third effect can dominate the first two effects, and the agent will prefer to spread reminders out rather than concentrate them close to the deadline.

### 1.6.2 Three-Period Example

To illustrate the SF’s solution with a complete menu of memory aids, we return to our 3-period example, where  $b = 1$ , and  $c \sim U[0, 2]$ . Let  $a^{ijk}$  denote the memory aid with  $a_1 = i, a_2 = j$ , and  $a_3 = k$ . For example,  $a^{011}$  denotes a

memory aid with reminders in the 2nd and 3rd periods.

Table 5 below summarizes the SF's willingness-to-pay for the 7 possible memory aids, for 4 sets of memory parameters:  
 $(p, \theta) \in \{(0.1, 0.1), (0.1, 0.9), (0.9, 0.1), (0.9, 0.9)\}$ .

	$p = 0.1$		$p = 0.9$	
	$\theta = 0.1$	$\theta = 0.9$	$\theta = 0.1$	$\theta = 0.9$
1 reminder				
$a^{100}$	0.2354	0.2086	0.0413	0.0122
$a^{010}$	0.2472	0.2110	0.0555	0.0143
$a^{001}$	0.2372	0.2111	0.0451	0.0140
2 reminders				
$a^{110}$	0.3695	0.3333	0.0652	0.0240
$a^{101}$	0.3704	0.3342	0.0662	0.0250
$a^{011}$	0.3709	0.3347	0.0667	0.0255
3 reminders				
$a^{111}$	0.4544	0.4182	0.0759	0.0347

Table 5: SF's WTP for Memory Aids,  $b = 1, c_t U[0, 2], T = 3$

Table 5 illustrates how the relative sizes of  $p$  and  $\theta$  can dictate whether the agent will prefer to spread out or backload reminders. Consider the relative WTPs for a single reminder. When  $p$  is low and  $\theta$  is high the impact of  $\theta$  on the probability of recall each period is minimized<sup>12</sup> and the agent strictly prefers later deadlines (see column 2). However, in all other scenarios, when either  $\theta$  is low, or  $\theta$  is high but  $p$  is also high, the agent prefers the reminder in the middle period. Note that, with memory aid  $a^{010}$ ,  $\theta$  is eliminated from

<sup>12</sup>This is true because  $\theta$  only appears in the expressions of expected utility when interacted with  $p$ .

the problem: the probability of recall in any period depends only on  $p$ , since the number of successive forgettings entering any period is always 0. While the expected utilities with the two other single-reminder aids are increasing in  $\theta$ , the expected utility with  $a^{010}$  is constant with respect to  $\theta$ . Thus, for a given  $p$ , lower values of  $\theta$  improve the value of  $a^{010}$  relative to  $a^{100}$  and  $a^{001}$ .

In contrast to SFs, NFs will never purchase memory aids. The reason is analogous to the explanation for why naive hyperbolic discounters do not use commitment devices. Since NFs expect to remember the task with probability 1 in all future periods, any memory aid with cost greater than 0 will be perceived as a strictly welfare-reducing choice. This leads directly to the theorem below:

**Theorem 9.** *Even with a complete menu of memory aids, a NF's welfare loss relative to that of an otherwise identical PR grows arbitrarily large as  $b \rightarrow \infty$ .*

*Proof of Theorem 9.* Trivial as well. Even when  $\theta = 1$ , the NF's utility from a task is at most  $p$  times the PR's utility of the same task. And as  $b \rightarrow \infty$ , holding the cost distribution fixed, we get that  $E_0(V_1(s^*(1, 1), 1, 1)) \rightarrow \infty$ . So,  $E_0(V_1(s^*(1, 1), 1, 1)) - E_0(V_1(s^*(\hat{p}, \hat{\theta}), p, \theta)) > (1-p)E_0(V_1(s^*(1, 1), 1, 1))$  grows arbitrarily large as well.  $\square$

Thus, incorporating memory aids into the model amplifies the cost of PM overconfidence. Without memory aids, overconfidence is welfare-reducing because agents set excessively low threshold costs. With memory aids, overconfidence is doubly costly – threshold costs are still too low, plus the agent abstains from purchasing welfare-improving memory aids.

Incorporating memory aids into the model provides an explanation for the experimental finding that subjects' retrospective and prospective memory performance may be negatively correlated [89]. Consider a group of subjects who are SFs, but with different true memory parameters. If memory aids are sufficiently costly, they will only be employed by agents with poor memories. Thus, we may observe that agents who have poorer memories (as tested through retrospective recall) may perform better on prospective memory tasks, but only because they are more likely to employ memory aids.

### 1.6.3 Present-Biased Preferences and Imperfect Memory

Earlier, we asserted that Shafir and Tversky's [75] result, that the probability of task completion could be decreasing in the length of deadline, cannot be derived from hyperbolic discounting alone. We will explicitly show this by incorporating time-discounting into our prospective memory model, and proving that any hyperbolic discounter (naive or sophisticated) with perfect memory will be made strictly better off by lengthening the deadline.

We choose a basic quasi-hyperbolic form of discounting, where the discount factor  $\beta$  between the present and all future periods is less than one, while the discount factor between all future periods  $\delta = 1$ . We set  $\delta = 1$  for analytical tractability, and maintain that choosing a  $\delta$  close to, but less than 1 will not meaningfully alter our results.

Furthermore, we assume, as in O'Donoghue and Rabin [62], that upon task execution, costs are incurred immediately, while benefits are delayed. Thus, the utility from executing a task in period  $t$  is now  $\beta b - c_t$  (as opposed to  $b - c_t$  in our baseline model).

Hyperbolic discounters determine their threshold costs each period in a similar manner to agents in our baseline model, with some important distinctions. First, note that with hyperbolic discounting in any period  $t$  for given memory parameters  $p, \theta$ , the value of a particular task which we denote  $V_t^H(s, p, \theta, \beta)$  is less than  $V_t(s, p, \theta)$ . In particular,

$$V_t^H(s, p, \theta, \beta) = \beta V_t(s, p, \theta).$$

Furthermore, for partially or fully naive hyperbolic agents,<sup>13</sup> anticipated threshold costs are greater than true threshold costs. Intuitively, naive hyperbolic agents believe that they will not discount the deferred benefit as greatly in the future, and thus predict threshold costs will be higher than they are. Define  $\hat{s}_t^H$  as the anticipated threshold cost in period  $t$ , from the perspective of any period prior to  $t$ . Then:

$$\hat{s}_t^H = \begin{cases} \hat{\beta} b & \text{if } t = T \\ \hat{\beta} b - E_t(V_{t+1}^H(\hat{s}^H, \hat{p}, \hat{\theta}, \hat{\beta})) & \text{if } 1 \leq t < T, \end{cases}$$

---

<sup>13</sup>Defined in [63] as agents who anticipate the value of  $\beta$  in future periods –  $\hat{\beta}$  – to be greater than its current value,  $\beta$ .

while

$$s_t^H = \begin{cases} \beta b & \text{if } t = T \\ \beta b - E_t(V_{t+1}^H(s^H, \hat{p}, \hat{\theta}, \hat{\beta})) & \text{if } 1 \leq t < T, \end{cases}$$

Note that a naive hyperbolic agent with  $\hat{\beta} = 1$  anticipates the same threshold costs as a SF in our model. That is, for  $\hat{\beta} = 1$ ,

$$\hat{s}_t^H = \begin{cases} b & \text{if } t = T \\ b - E_t(V_{t+1}^H(\hat{s}^H, \hat{p}, \hat{\theta}, 1)) & \text{if } 1 \leq t < T, \end{cases}$$

which is exactly the set of threshold costs for a SF with PM beliefs  $\hat{p}, \hat{\theta}$  (see equation 1). Given this, the naive hyperbolic's actual set of thresholds is:

$$s_t^H = \begin{cases} \beta b & \text{if } t = T \\ \beta [b - E_t(V_{t+1}^H(\hat{s}^H, \hat{p}, \hat{\theta}, 1))] & \text{if } 1 \leq t < T, \end{cases}$$

or simply  $\beta s_t(\hat{p}, \hat{\theta})$ , which is the fraction  $\beta$  of the optimal cost threshold at  $t$  if the “true” memory parameters were  $\hat{p}, \hat{\theta}$ .

Note, for a sophisticated hyperbolic agent with  $\hat{\beta} = \beta$ ,  $s_t^H = s_t^H$  for all  $t \in \{1, \dots, T\}$ . For a given  $\beta$ , the sophisticated hyperbolic agent's anticipated and executed cost thresholds are below the naive hyperbolic agent's anticipated thresholds, and above the naive hyperbolic agent's actual thresholds.

We now assess the welfare effect of lengthening the deadline on perfect-remembering hyperbolic discounters:

**Theorem 10.** *For any hyperbolic discounting agent with  $p = \hat{p} = \theta = \hat{\theta} = 1$ , the probability of task completion and expected utility from task completion are strictly increasing in the time allotted.*

The logic behind this theorem is the same as for Theorem 5 and the proof is similar.

*Proof of Theorem 10.* The proof of this theorem is virtually identical to that of Theorem 5. Any perfectly-remembering agent, whether hyperbolic discounter



or not, faces the exact same problem in the last  $T$  periods of a  $T + \delta$  time frame as in a  $T$  period time frame and thus over the last  $T$  periods executes the exact same strategy with the same expected utility conditional on arriving in that period. In the  $\delta$  periods prior to this replication, a hyperbolic discounter will only complete the task if the cost draw is low enough to increase overall expected utility. Thus, adding periods strictly increases the expected utility of any PR in this context, regardless of the time-discounting parameters. The increasing probability of task execution is analogous.  $\square$

While our length of deadline result cannot be generated solely from hyperbolic time-preferences, we do not mean to imply that hyperbolic time preferences do not importantly influence agent behavior in our model. Rather, we believe there are positive interaction effects between naivete with respect to fallible memory and procrastination. In particular, for agents with imperfect memory and hyperbolic time preferences, the cost of fallible memory is increasing in the tendency to procrastinate (as captured by  $\beta$ ), and vice versa. Table 6 shows the ex ante probability of task execution and the non-discounted expected utility over a range of  $p$  and  $\beta$  for naive-forgetting, naive-hyperbolic agents, for our 3-period model with  $b = 1, c \sim U[0, 2]$ , and  $\theta = 0.2$ :

$p/\beta$	Probability of Task Completion				Expected Utility			
	0.2	0.5	0.8	1.0	0.2	0.5	0.8	1.0
0.2	0.019	0.047	0.073	0.089	0.018	0.039	0.053	0.060
0.5	0.067	0.156	0.233	0.278	0.062	0.127	0.166	0.180
0.8	0.143	0.324	0.469	0.546	0.132	0.262	0.328	0.345
1.0	0.215	0.478	0.676	0.777	0.198	0.384	0.468	0.483

Table 6: Probability of Task Completion and Ex Ante (Non-Discounted) Expected Utility for NF Naive-Hyperbolics,  $b = 1, c_t \sim U[0, 2], T = 3, \theta = 0.2$

Note, for example, that for  $\beta = 0.8$ , the probability of completion goes from 0.469 to 0.073 as  $p$  goes from 0.8 to 0.2, a reduction of 84%. For  $\beta = 0.2$ , the probability goes from 0.143 to 0.019, a reduction of 87%. Similarly, fix  $p = 0.8$  – the percentage reduction in the probability of completion for  $\beta = 0.2$  vs  $\beta = 0.8$  is 70%, while for  $p = 0.2$  the reduction is 74%. Similar interactions hold for expected utility. These results suggest that the mixed partials with respect to  $p$  and  $\beta$  of both  $\log Q$  and  $\log V_0$  are positive. In words,

the percentage reduction in the ex ante probability of task execution and non-discounted expected utility as  $p$  is lowered are decreasing in  $\beta$ , and decreasing in  $p$  as  $\beta$  is lowered.

In our examples, we found that this same pattern – that the cost of fallible memory is increasing in the tendency to procrastinate, and vice versa – also holds true for naive-forgetting, sophisticated-hyperbolic agents; sophisticated-forgetting, naive-hyperbolic agents; and sophisticated-forgetting, sophisticated-hyperbolic agents (results not shown).

Note however, that the interaction of fallible memory and procrastination is dependent on the agent’s relative sophistication with respect to her memory parameters and time preferences. As discussed in Ericson [28], for a sophisticated forgetter, fallible memory may effectively act as a commitment device against procrastination, elucidating that the “now or later” choice is actually “now or never.”

## 1.7 Prospective Memory Failures in the Marketplace

### 1.7.1 Mail-In Rebates

Many consumer goods companies (particularly electronics manufacturers) offer mail-in rebates, ranging in value from less than \$1 to several hundred dollars. To claim the rebate, purchasers generally must fill out an information card and mail it (with proof of purchase) to a provided address within some time frame after the purchase. This time frame generally varies from 4 weeks to 3 months. Experts estimate that the rebate market has grown from under \$1 billion to \$4-\$10 billion in the past decade, and that the number of rebates offered are in the hundreds of millions.<sup>14</sup> However, typically only 5% of outstanding rebates are ever claimed by consumers, and while claim rates rise with the value of the rebate, redemption rates are typically less than 40% even for “big-ticket” electronics items with rebate values greater than \$20.<sup>15</sup>

While the traditional explanation for rebates is that they are a form of

---

<sup>14</sup>See [38]

<sup>15</sup>See [80]

price discrimination by retailers, an increasing body of evidence suggests that a significant amount of rebate non-redemption (or “breakage”) comes from consumers who initially planned on redemption – a dynamic not captured by the price discrimination model. In addition to the experimental studies cited earlier ([77] and [27]), Silk and Janiszewski [78], in a survey of 35 marketing managers at firms that use rebates, find that those managers estimate that nearly two-thirds of breakage is attributable to customers that had intended to redeem the rebate, but for some reason or another, failed to do so. One possible reason for this breakage is that firms are exploiting naive forgetting by consumers. Naive forgetters fully internalize rebate discounts in making purchase decisions, failing to account for the prospective memory lapses that often result in non-submission. Firms may have recognized that through this naive forgetting, rebates increase the effective downward price elasticity of demand, and consistent with profit-maximization, have made greater use of them.

There are other explanations for rebates beyond price discrimination and naive forgetting, both on the consumer and firm sides. Consumers may be naive not about their forgetfulness, but rather their present-biased preferences [62]. That is, consumers may purchase goods expecting to complete the rebate process, not anticipating their tendency to procrastinate on such onerous tasks in the future. This explanation may hold true for lesser value rebates, but is calibrationally inconsistent with low claim rates for rebates worth \$100 or more. Chen, Moorthy, and Zhang [13] develop a theory of rebates for perfectly rational consumers, where rebates facilitate “utility arbitrage” through state-dependent pricing. However, such a rational explanation is inconsistent with anecdotal evidence of consumers’ ex post regret and frustration over rebate programs. The New York Times reported that rebate complaints to the Better Business Bureau increased nearly three-fold from 2001 to 2005 (from 964 to 2,715), and US News and World Report more recently documented that the complaint volume cleared 4,500 in 2007.<sup>16</sup> Furthermore, due to pressure from consumers and advocacy groups, legislation has been proposed in California and Texas and passed in New York to regulate rebate practices. The New York law addresses rebate form availability, establishes a 2-week minimum redemption period, requires rebate rewards to be paid within 60 days of submissions, and restricts retailers from misleading advertising of post-rebate

---

<sup>16</sup>See [85] and [65]

prices.<sup>17</sup>

On the firm side, rebates have been justified on grounds other than price discrimination, such as collecting consumer data. While these explanations may have some validity, rebates would not likely be as commonly offered if claim rates were closer to 100%. Larissa Hall, the vice president of marketing at Buy.com, stated that “the reason we can offer rebates is because not everybody will redeem them” [16]. Analysts tracking consumer goods companies have claimed that “rebates are a good business plan only when consumers fail to claim them,”<sup>18</sup> and fulfillment companies often market their services to manufacturers by touting the low redemption rates for promotions they have administered.

One prediction of our model is that firms should lengthen rebate deadlines, or even restrict rebate submissions until after some grace period, to induce more forgetting by naive consumers. While we have no systematic evidence to this effect, we have observed instances where new rebate practices do seem geared towards further capitalizing on forgetting by consumers. For instance, a number of fulfillment companies are now offering manufacturers and retailers “deferred” rebate programs, where consumers have to file initial paperwork soon after the purchase, but then must file an additional claim as much as a full year later. These “deferred” rebate programs generally offer substantially greater discounts than regular rebates, but these discounts are more than offset by the lower claim rates over the lengthened rebate process. Additionally, several on-line retailers have offered free-after-rebate deals – essentially a 100% rebate – while initially pricing goods above the suggested retail price. Indeed, a number of Internet retailers made such promotions the core of their business strategy.<sup>19</sup> As with deferred rebate plans, free-after-rebate deals can be explained as an attempt by manufacturers and retailers to further capitalize on naive forgetting by consumers. Edwards [22], in commenting on this increasingly evident strategy by firms, notes that one potential issue with the New York rebate law is that “it does not set any maximum time period....[and] some scholarship has suggested that longer deadlines can lead to fewer rebate redemptions due to increased chances for consumer procrastination and forgetfulness.”

---

<sup>17</sup>See [22]

<sup>18</sup>See [55]

<sup>19</sup>See [13] for a more detailed discussion of this sales tactic.

## 1.7.2 Free Trial Offers

A growing practice among firms with subscription-based products is to offer consumers “risk-free” trials. Consumers are not billed and may cancel at any time during the trial. However, if the trial offer is not canceled before its expiration, it is automatically renewed at the full subscription price. An agent with imperfect memory may learn during the trial that she would not be willing to renew the subscription, but simply forget to cancel before the deadline, and incur a loss as a result. Zaidi [95], in a field experiment with a subscription-based website, finds that varying the strength of the reminder that subscribers receive on their credit card bill has a significant effect on cancellation rates – evidence that forgetting does play a role in consumer behavior with respect to subscription-based (or more broadly, “negative option”) goods.<sup>20</sup> Overconfidence in prospective memory could further exacerbate the welfare loss from unwanted renewals. We collected data on a free trials run by a subscription-based website that appears consistent with the key prediction of our model incorporating memory overconfidence – that task incompleteness caused by naive forgetting may be increasing in the time allotted before the deadline. The website randomly e-mailed prospective subscribers one of two trial offers – one for 3 days, and one for 7 days. Trial subscriptions were automatically renewed on a monthly basis at the end of the trial. Post-trial retention was 28% for the 3-day trial group, and 41% for the 7-day trial group. The difference in retention rates is significant at 99% confidence.<sup>21</sup>

The proliferation of credit cards over the past 20 years has greatly facilitated the recurrent billing process and opened up virtually the entire US con-

---

<sup>20</sup> In the field test, new members are randomly assigned to standard-pricing (e.g., \$19.95) and “camouflage” pricing (e.g., \$20.13 or \$19.83) groups. Assuming that consumers’ credit card statements serve as a reminder of the subscription and of the option to cancel, the hypothesis was that the camouflage pricing group would have lower cancellation rates, since the camouflage amounts tend to blend in with other charges on a credit card bill (e.g., gas, groceries), while standard amounts are more likely to stand out on a bill, and be noticed by consumers who are in effect “reminded” of their subscription. The field study results supported the hypothesis as camouflage pricing lowered the odds of cancellation by roughly 10% over the first 4 months of membership.

<sup>21</sup> The free trials were offered to consumers on the website’s mailing list, most of whom were past subscribers. To the extent that past subscribers already have a sense for the quality of the website, the learning/information explanation for the existence of free trial offers (discussed later in this section) is less likely to explain the disparity in retention rates.

sumer population to be targeted by negative option marketers.<sup>22</sup> Publishers and other companies selling subscription-based products and services have not only unilaterally marketed free trials, but partnered with third parties in marketing free trials as a “bonus” item. For instance, most major US airlines allow customers to redeem miles for short, 3-6 month magazine subscriptions that are automatically renewed at the regular subscription rate. Retailers (particularly online) often offer these subscriptions as a “free bonus” with any purchase.<sup>23</sup> A relatively recent phenomenon is “incentive” marketing, where intermediary companies offer consumers gifts for signing up for multiple trial offers at once. These companies are paid \$40-\$60 per free trial signup, leaving them a comfortable profit margin when rewarding 5-10 free trial signups with a \$100-\$200 gift.<sup>24</sup> One such company, Gratis Internet, grossed over \$20M in 2005, and reported giving away over 20,000 iPods to consumers as rewards for their signups.<sup>25</sup>

Anecdotal evidence suggests that firms are increasing the length of their free trials, possibly to increase naive forgetting by consumers. Many weekly magazines (such as Sports Illustrated) have increased the standard length of their free trials from 4 issues to 3 months. In the past 5 years, AOL has increased the length of their dial-up service free trial offer from 30 to 50 days. In addition, a new class of health-related products – weight-loss supplements, smoking-cessation aids, even teeth whiteners – has emerged via direct marketing channels such as the internet and infomercial programming. Their manufacturers offer extended free or reduced-price trials – of lengths running in months rather than days or weeks – ostensibly to give their new customers as much time as possible to test and experience the benefits of their prod-

---

<sup>22</sup>Before credit cards were widespread, customers generally paid for subscriptions by check or money order. According to a publishing executive we interviewed on this subject, payment rates for free trial customers who were billed for automatic extensions through the mail were extremely low, reducing the attractiveness of this strategy.

<sup>23</sup>One example of this is a promotion run by the Sam Goody music store chain, where customers making a purchase were offered a three-month “free trial” subscription to Entertainment Weekly magazine. Customers filed a class-action suit against the retailer in August 2003, claiming they were not informed that their billing information would be passed on to the publisher and that their credit cards would automatically be charged for a full subscription at the end of the trial period.

<sup>24</sup>See [87]

<sup>25</sup>See Inc.com company profile at [www.inc.com/inc5000/2007/company-profile.html?id=2005018](http://www.inc.com/inc5000/2007/company-profile.html?id=2005018)

ucts, but effectively as a way to more fully capitalize on consumer forgetting. This extended-trial tactic has garnered such artificially high take-up rates that it has elicited a high volume of consumer complaints to the Better Business Bureau, Federal Trade Commission, and credit card companies. “Acai Supplements and Other ‘Free’ Trial Offers” ranked #1 on the BBB’s List of the Top 10 Scams and Rip-Offs in 2009.<sup>26</sup> The BBB reported that one company selling acai berry supplements received more complaints in 2009 than the entire airline industry combined.<sup>27</sup> Visa recently removed 100 companies that were marketing “deceptive free trial offers” from its authorized vendor list, after a rash of customer complaints and chargeback requests.<sup>28</sup> The extent of ex post consumer regret and frustration over these offers is not only suggestive that consumers are incurring a loss of welfare from this marketing strategy, but also runs counter to the rational explanation for free trials – that they simply offer consumers a free or low-cost opportunity to learn about their product before making a greater investment. It is hard to imagine consumers generating this volume of complaints just from learning about their preferences.

While the welfare loss from any one free trial transaction may be small, the aggregate loss generated by this exploitation of naive forgetting may be quite significant. Magazine industry revenues topped \$20 billion in 2009,<sup>29</sup> and many other products and services use negative option models, including telecommunications providers, music and book clubs, and information services (such as credit report agencies). Visa, in a survey of 1,000 cardholders in September 2009, found that nearly 30% had been “duped” by some form of negative option marketing. With over 575 million credit cards and 500 million

---

<sup>26</sup>See full BBB Bulletin at [www.bbb.org/us/article/](http://www.bbb.org/us/article/bbb-lists-top-10-scams-and-rip-offs-of-2009-14436)  
[bbb-lists-top-10-scams-and-rip-offs-of-2009-14436](http://bbb-lists-top-10-scams-and-rip-offs-of-2009-14436)

<sup>27</sup>See [90]

<sup>28</sup>See full notice to Visa customers at [usa.visa.com/personal/security/learn-the-facts/deceptive-marketing.html?ep=v\\_sym\\_negativeoption](http://usa.visa.com/personal/security/learn-the-facts/deceptive-marketing.html?ep=v_sym_negativeoption). Chargebacks are the main drawback of this marketing tactic, according to a publishing executive that we interviewed on the subject of automatic renewals. Credit card chargebacks occur when customers call their credit card issuer directly to dispute a charge. The card issuer then charges the merchant back for the amount, unless the merchant files a response. For most merchants (including magazine publishers), the administrative cost of filing responses is too great to contest each chargeback, and thus chargebacks usually stand. If a merchant’s chargebacks exceed some threshold (usually around 2% of all transactions) in a billing period, the merchant is assessed a large penalty (on the order of \$100,000).

<sup>29</sup>According to year-end numbers published by the Publishers Information Bureau at [www.magazine.org/advertising/revenue/by\\_mag\\_title\\_ytd/pib-4q-2009.aspx](http://www.magazine.org/advertising/revenue/by_mag_title_ytd/pib-4q-2009.aspx)

debit cards issued to 175 million cardholders in the US by year-end 2009,<sup>30</sup> an extrapolation of that 30% to the broader credit-card-carrying consumer population would suggest that the welfare loss from consumers' prospective memory failures is pervasive and considerable.

## 1.8 Areas for Future Research

As discussed in Sections 1.2 and 1.4, empirical evidence on prospective memory beliefs and performance is somewhat limited. The experimental results in Silk [77] and Ericson [27] are strongly suggestive that naive forgetting impacts consumer behavior beyond the laboratory. Rebates and free trial offers are two ways that firms appear to exploit naive forgetting to consumers' detriment, and more field work in these areas would be valuable in calibrating the overall welfare loss from these deviations from the standard model. Ideally, field tests could be arranged in a "revealed-preference" design, giving subjects the choice between immediate payoffs and "rebate coupons" of varying values and deadlines to enable a precise identification of consumers' memory beliefs and true parameters (in the same vein as Ericson's experimental design). Another interesting treatment would be to allow consumers to choose their deadline length, as in Ariely and Wertenbroch [6].<sup>31</sup> Field evidence on rebates and free trials may also shed greater light on the equilibrium effects of fallible memory that may be obscured in experimental settings, as memory aids and other commitment-type devices may be effectively used to limit welfare losses in "real-world" settings. For instance, consumer advocacy groups are now touting the use of disposable, "virtual credit cards" – numbers that are issued for a single-use only, and then invalidated – for the purposes of avoiding unwanted takeup with negative option goods.<sup>32</sup>

From a theory standpoint, there are potentially interesting extensions of our model that we have not yet examined. In order to compare and con-

---

<sup>30</sup>See link to Visa's notice to customers references in footnote 26.

<sup>31</sup>This proposed study would be different in that the task – completing a rebate – would have greater applicability to our PM model. Their task – completing a class project – may have too many natural reminders to be relevant in our context.

<sup>32</sup>This service is offered by most large credit card companies, as well as payment services such as Paypal.



trast our findings to the present-biased preferences literature, we have thus far focused on examining onerous tasks – that is, ones with upfront costs and deferred benefits. We have followed the literature in assuming an environment with stochastic costs and some fixed benefit to task execution. However, our framework is more generally applicable to environments with time-varying costs and benefits (whether stochastic or deterministic). In such environments, it may be interesting to examine the bounds on losses (relative to the first-best) for different types of agents.

Another important theoretical consideration concerns potential learning about prospective memory. This learning may occur across tasks (e.g., if I forget to submit one rebate, I submit my next rebate right away), or more relevantly to our current framework, within tasks. In particular, it may be that NFs update their beliefs such that after a string of recall failures, they revise beliefs downward and correspondingly increase their threshold costs. For instance, in everyday life, often times if we have forgotten to perform a task for a long while and it suddenly pops into our head, we may think “better do it now before I forget again.” However, if, by luck of the draw, the same task continually comes to our attention, we are much more likely to think we can defer the task and bank on future recall.

The model of prospective memory we presented in the first chapter was defined entirely at the level of the individual. The intertemporal, intrapersonal dynamics that arose from the multi-period nature of the choice framework we defined created game-theoretic considerations when solving the model, but nonetheless, the model is single-actor, very much in the spirit of the Psychology and Economics field of economic theory. The model of left tail dependence presented in the next chapter lies at the opposite end of the spectrum, in that it models dependence in financial markets from a purely top-down perspective, without delving into the motivations or psychologies of market participants. Despite the different methodological approaches of the two chapters, they share a goal of modeling empirical phenomena that standard economic models have difficulty incorporating. In the first chapter, the phenomena were low rates of completion of worthwhile tasks even in the face of deadlines, and the misprediction of these low rates of completion. In the second chapter, the phenomenon is the empirical observation that left tail events occur simultaneously across financial markets to a degree that is surprising given the lack of extreme dependence in the bulk of the multivariate distributions. We model this dependence with an innovative new copula that permits arbitrarily extreme dependence in the tails, while allowing for a general dependence structure in the bulk.

## 2 Left Tail Dependence

### 2.1 Introduction

Modeling the bulk and the tails of a multivariate distribution simultaneously requires a balancing act that is difficult if not impossible to achieve with the well-known copulas. In this paper we construct a multivariate density which resembles an elliptical distribution in its bulk, but has the following two properties in its tails:

1. Association that is more extreme in the tails than in the bulk of the distribution
2. Association that is asymmetric between the upper and lower tails

We achieve this by introducing a new “Cube copula”, that can accommodate arbitrarily precise modeling of the joint tails. We then describe two methods to break the link between the tails and the bulk of the distribution, copula mixing and copula nesting, and we illustrate with an application to hedge fund returns.

Copulas have become a key addition to the financial modeler’s toolkit in the past decade. Formally, a copula is a multivariate cumulative distribution function defined on the unit cube such that its marginals are uniform. A key theorem, due to Sklar [79], lays out the use of such a function – for any set of marginal distributions defined on the reals, and any joint distribution function, there exists a copula that reproduces the joint distribution when applied to the marginal distribution functions. Modulo continuity conditions on the marginals, this copula is unique. The reason for the appeal to financial modeling is that researchers often have better information about marginal distributions than joint distributions, and a copula approach lets them fully use this information, and then choose from a small menu of well-known copulas to splice them into a joint distribution.

The set of well-known copulas leaves much to be desired when modeling financial returns. A brief, stylized history of univariate return modeling

is instructive. Early modeling of financial (log) returns relied heavily on the Gaussian distribution, primarily for its analytical tractability. However, returns typically exhibit kurtosis much greater than that of the Gaussian, and so modelers expanded into fatter tailed distributions, such as the Student's-t and Generalized Extreme Value. In the multivariate case, the Gaussian copula played a similarly tractable role in early applications. The Gaussian copula is defined (via Sklar) as that copula which composes univariate Gaussians into a multivariate Gaussian. Just as in the univariate case, the Gaussian copula imposes a particular structure in the joint tails of multivariate distributions that is often empirically violated. Specifically, and we define this concept rigorously in section 2.2.4, the Gaussian copula requires that variables become asymptotically independent in the tails, while in practice, dependence even in extreme tail events often remains strong.

The “fix” has often been to simply move to a copula with a fatter joint-tail, such as the Student's-t copula, which is that copula that composes Student's-t marginals into a multivariate Student's-t. Other popular copulas with non-zero asymptotic tail dependence are the Archimedean copulas, which encompasses the Clayton, Frank, and Gumbel copulas. However, in making this move, one loses control of the ability to model both the bulk of the multivariate distribution and the joint tails. For example, the bivariate Student's-t copula has two parameters,  $\eta$ , the degrees of freedom, and correlation  $\rho$ . The amount of left tail dependence is a decreasing function of  $\eta$  and an increasing function of  $\rho$ . Thus  $\rho$  and  $\eta$  can serve as tuning parameters for tail dependence. However, neither of these parameters changes solely tail dependence. The  $\rho$  parameter is in fact the correlation of the bivariate distribution in the case of Student's-t marginals, so that a side effect of increasing tail dependence via increasing  $\rho$  is a stronger dependence in the bulk of the distribution. Frequently in applications we will have an estimate of correlations in the bulk of the distribution, and want to increase tail dependence while holding fixed the correlation in the bulk. The  $\eta$  parameter is less intuitive, but has equally unattractive properties upon scaling. First,  $\eta$  has limited ability to generate extreme tail dependence. Figure 3 plots the tail dependence as a function of  $\eta$  for several values of  $\rho$ .

For example, with  $\rho = 0.3$  the maximum achievable tail dependence is 0.41. Furthermore, very low values of  $\eta$  generate behavior that most would consider odd in the upper left and lower right quadrants of the multivariate.

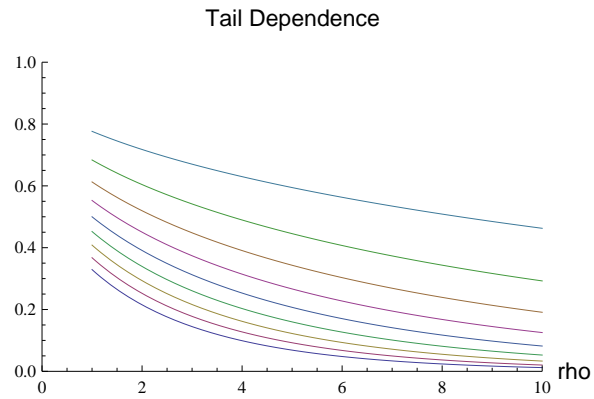


Figure 3: Tail dependence  $\lambda_L$  as a function of  $\eta$  for several  $\rho = 0.1 \dots 0.9$ .

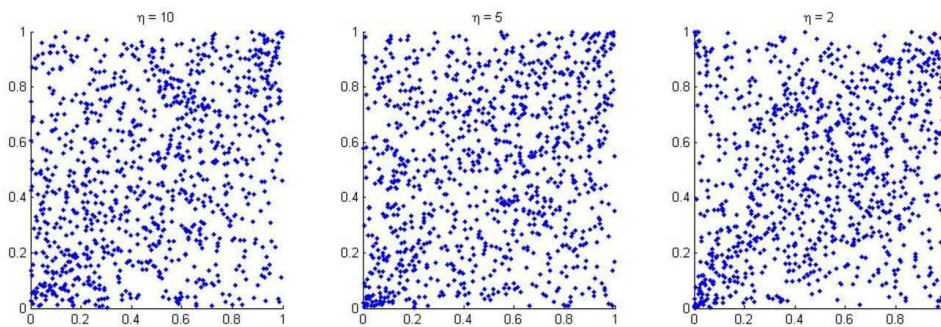


Figure 4: the  $t$  copula with  $\rho = 0.3$  for  $\eta \in \{10, 5, 2\}$

Fig. 4 plots a Monte Carlo simulation of the  $t$  copula with  $\rho = 0.3$  for  $\eta \in \{10, 5, 2\}$ . Note the extreme right tail outliers in univariate  $Y$  that are associated with left tail outliers in univariate  $X$  with increasing frequency as  $\eta$  decreases. The intuition for this behavior is that the  $t$  copula needs to enforce correlation  $\rho$ , and so must balance out what we call the “double-tail” observations with “anti-double-tail” points.

## 2.2 The Cube Copula

### 2.2.1 Construction in n-dimensions

**Definition 1.** Let  $X$  be a random variable on the sample space  $\Omega_n = [0, 1]^n$ . Let  $\mathbf{a}$  be an element of  $\Omega_n$ . If  $X_i \leq a_i$  then we say that  $X$  experiences an  $i$ -th tail event (with respect to  $\mathbf{a}$ ).

For an arbitrary set of indices  $I \subseteq \{1, \dots, n\}$ , define the set

$$T_I = \{x \in \Omega_n \mid x_j \leq a_j \text{ if and only if } j \in I\}.$$

Furthermore when  $I$  has cardinality  $|I| = k$ , we refer to  $T_I$  as a  $k$ -tail region.

In other words, a  $k$ -tail region is a subset of the sample space with exactly  $k$  of the variables experiencing tail events simultaneously. For any fixed  $k$ , the number of  $k$ -tail regions equals  $\binom{n}{k}$ .

Let  $\tau_{k,n}$  denote the union of all  $k$ -tail regions in  $\Omega_n$ . The sets  $\tau_{k,n}$  for various  $k$  form a partition of  $\Omega_n$ ; that is,

$$\Omega_n = \bigcup_{k=0}^n \tau_{k,n}, \quad \tau_{k,n} \cap \tau_{\ell,n} = \emptyset \text{ if } k \neq \ell. \quad (2)$$

Figure 5 illustrates the 2-tail and 3-tail regions for  $n = 3$  and  $\mathbf{a} = (0.1, 0.1, 0.1)$

**Theorem 11.** Consider a real vector  $\mathbf{q} = (q_0, \dots, q_n) \in \mathbb{R}^{n+1}$  and the corresponding density on  $\Omega_n$ :

$$p_c(x) = q_k \text{ when } x \in \tau_{k,n}. \quad (3)$$

Then  $p_c$  is a copula density if and only if conditions (a) and (b) below are met:

(a) Total Density Condition

Defining  $v_{k,n}[\mathbf{a}] := \text{vol}(\tau_{k,n})$ , one has

$$1 = \sum_k q_k v_{k,n}[\mathbf{a}], \quad (4)$$

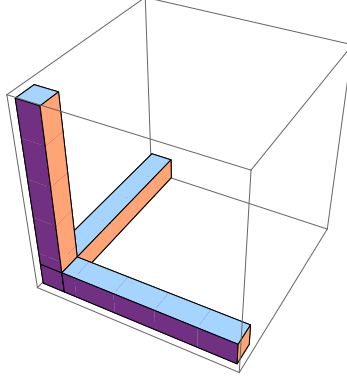


Figure 5: The 2-tail and 3-tail regions for  $n = 3$

(b) Unit Marginals Condition

Defining  $\hat{\mathbf{a}}^{(j)} = (a_1, a_2, \dots, a_{j-1}, a_{j+1}, \dots, a_n)$ , one has for all  $j$ :

$$M_{j,1} := \sum_{i=1}^n q_i v_{i-1, n-1}[\hat{\mathbf{a}}^{(j)}] = 1 \quad (5)$$

$$M_{j,2} := \sum_{i=1}^n q_{i-1} v_{i-1, n-1}[\hat{\mathbf{a}}^{(j)}] = 1. \quad (6)$$

Furthermore, one has

$$v_{k,n}[\mathbf{a}] = \sum_{|I|=k} \text{vol}(T_I) = \sum_{|I|=k} \left[ \prod_{i \in I} a_i \prod_{j \in I^c} (1 - a_j) \right]. \quad (7)$$

where the sum is over the  $\binom{n}{k}$  subsets  $I \subseteq \{1, 2, \dots, n\}$  with length  $k$ , and  $I^c$  is the complement of  $I$ .

*Proof.* In order that (3) be a probability density, we must have

$$\int_{\Omega_n} p_c(x) dx = 1 = \sum_k v_{k,n}[\mathbf{a}] q_k, \quad \text{where } v_{k,n}[\mathbf{a}] := \text{vol}(\tau_{k,n}). \quad (8)$$

This is condition (a) in the theorem. To prove (7), and thus derive an explicit formula for (4), we must calculate  $v_{k,n}[\mathbf{a}]$ . Note that  $\tau_{k,n}$  is the union of  $\binom{n}{k}$

connected components, each a rectangular prism. Let  $I \subseteq \{1, 2, \dots, n\}$ ; then for  $k = |I|$  the region  $T_I$  defined above is one of the  $\binom{n}{k}$  components of  $\tau_{k,n}$ . The volume of such a region is the product of its side lengths, and the regions are disjoint as noted above in eq. (2). This establishes (7).<sup>33</sup>

In order that  $p_c(x)$  forms a copula density, we require the condition of uniform marginals:

$$m_j(x_j) = \int_{[0,1]^{n-1}} p_c(x) \prod_{i \neq j} dx_i = 1. \quad (9)$$

We claim that each  $m_j$  is a step function on  $[0, 1]$ , and more specifically can be written in the form

$$m_j(x_j) = \begin{cases} M_{j,1} & \text{if } x_j \leq a_j, \\ M_{j,2} & \text{if } x_j > a_j \end{cases}. \quad (10)$$

where  $M$  is an  $n \times 2$  matrix that is a polynomial function of  $\mathbf{q}$  and  $\hat{\mathbf{a}}^{(j)}$ . We now focus on establishing the representation (10) by calculating the required integrals explicitly.

The marginal  $m_j(x_j)$  is given by the integral of  $p_c$  over the lower-dimensional “slice”

$$S(x_j) = \{y \in \Omega_n : y_j = x_j\}$$

with respect to  $n - 1$  dimensional volume, i.e.

$$m_j(x_j) = \int_{S(x_j)} p_c(z) \prod_{i \neq j} dz_i. \quad (11)$$

Since  $p_c(x)$  is a step function, (11) can be written as a finite sum of density times volume; it remains to determine the explicit form of this sum.

The slice  $S(x_j)$  is isomorphic  $\Omega_{n-1}$  and hence has the same combinatorial structure as the original problem in one lower dimension. For each  $k = 0, 1, \dots, n - 1$  the slice  $S(x_j)$  has  $\binom{n-1}{k}$  connected components which play the role of  $k$ -tails in  $\Omega_{n-1}$ . The side lengths for the resulting partition of  $\Omega_{n-1}$  are

---

<sup>33</sup>In the special case  $\mathbf{a} = (a, a, \dots, a)$ , this simplifies to  $v_{k,n} = \binom{n}{k} a^k (1-a)^{n-k}$ .



determined by the truncated vector  $\hat{\mathbf{a}}^{(j)}$  defined in part (b) of the theorem. Hence our strategy is to use a form of induction on  $n$ .

First consider the case  $x_j \leq a_j$ ; then the  $\mathbf{q}$ -vector relevant for calculating the  $(n - 1)$  dimensional density on the slice is  $(q_1, q_2, \dots, q_n)$ , because  $S(x_j)$  doesn't intersect the 0-tail in  $\Omega_n$ . The set of points in  $S(x_j)$  where  $p_c = q_i$  has the structure of an  $(i - 1)$  tail region in  $\Omega_{n-1}$ , and is a subset of an  $i$ -tail in  $\Omega_n$ . Hence

$$M_{j,1} = \sum_{i=1}^n q_i v_{i-1, n-1}[\hat{\mathbf{a}}^{(j)}]. \quad (12)$$

For the case  $x_j > a_j$  the logic is the same, but instead of working with  $(q_1, q_2, \dots, q_n)$  we have to work with  $(q_0, q_1, \dots, q_{n-1})$  because  $S(x_j)$  intersects the 0-tail but not the  $n$ -tail in  $\Omega_n$ . In this case the set of points in  $S(x_j)$  where  $p_c = q_i$  has the structure is a subset of an  $(i - 1)$ -tail in  $\Omega_n$ . Hence we shift the index on  $q$  in (12) to yield

$$M_{j,2} = \sum_{i=1}^n q_{i-1} v_{i-1, n-1}[\hat{\mathbf{a}}^{(j)}]. \quad (13)$$

With this, we establish that formulas (5) and (6) are correct, and complete the proof.  $\square$

Note also that for any  $\mathbf{a}$  and for any  $n$  there is always at least one  $\mathbf{q}$  that trivially defines a Cube copula, namely  $\mathbf{q} = (1, 1, \dots, 1)$ . In fact we conjecture that there are always infinitely many such consistent  $\mathbf{q}$ ; the argument involves the degrees of freedom allowed in the Unit Marginals Condition.

For the remainder of the paper we assume that  $\mathbf{a} = (a, a, \dots, a)$ ; our results below hold more generally but this assumption simplifies notation considerably. For instance, with this assumption, the Total Density Condition and Unit Marginals Conditions of Theorem 11 simplify to just three equations:

(a) *Total Density Condition*

$$1 = \sum_k v_{k,n} q_k \text{ where } v_{k,n} = \binom{n}{k} a^k (1-a)^{n-k} \quad (14)$$

(b) *Unit Marginals Condition*

$$1 = \sum_{i=1}^n q_i v_{i-1, n-1} = \sum_{i=1}^n q_{i-1} v_{i-1, n-1}. \quad (15)$$

### 2.2.2 Examples: 2 and 3 dimensions

In  $n = 2$  dimensions eq. (7) implies

$$v = (v_{0,2}, v_{1,2}, v_{2,2}) = ((1-a)^2, 2a(1-a), a^2).$$

We then have the copula conditions as in Theorem 11:

$$\begin{aligned} (1-a)^2 q_0 + 2a(1-a)q_1 + a^2 q_2 &= 1 \\ (1-a)q_1 + aq_2 &= 1 \\ (1-a)q_0 + aq_1 &= 1 \end{aligned}$$

These equations can of course be solved explicitly; we choose to express the solution in terms of  $q_2$ , the density in the double-tail region:

$$q_0 = \frac{1-2a+a^2 q_2}{(a-1)^2}, \quad q_1 = \frac{aq_2-1}{a-1}.$$

Positivity of  $q_0, q_1$  give the constraint that  $(2a-1)a^{-2} \leq q_2 \leq a^{-1}$ . This implies  $0 \leq q_0 \leq (1-a)^{-1}$  and  $0 \leq q_1 \leq a^{-1}$ . These constraints are useful in maximum-likelihood optimization to guide the optimizer and ensure that it doesn't go outside the valid parameter space.

In  $n = 3$  dimensions eq. (7) implies

$$v = (v_{0,3}, \dots, v_{3,3}) = ((1-a)^3, 3a(1-a)^2, 3a^2(1-a), a^3).$$

Again as in Theorem 11, the conditions for a copula are:

$$\begin{aligned} (1-a)^3 q_0 + 3a(1-a)^2 q_1 + 3a^2(1-a)q_2 + a^3 q_3 &= 1 \\ (1-a)^2 q_1 + 2a(1-a)q_2 + a^2 q_3 &= 1 \\ (1-a)^2 q_0 + 2a(1-a)q_1 + a^2 q_2 &= 1 \end{aligned}$$

In any number  $n$  of dimensions, we can represent this system as  $Aq = \mathbf{1}$  where  $A$  is a  $3 \times n + 1$  matrix. Hence in  $n$  dimensions there is a space of solutions (copulas) which is of dimension  $n - 2 + p$ , where  $p$  is the dimension of the null space of  $A$ . We showed above that for  $n = 2$ , we have  $p = 1$  and the solution space is one-dimensional.

### 2.2.3 Properties of the Cube Copula

Above we referred to “tail dependence” informally, but there is a natural definition that is standard within the copula literature. We provide this definition, and compute the tail dependence of our Cube copula. This tail dependence, and its finite analog, serve as parameters that we can estimate/calibrate when applying the Cube copula to data. In addition, we consider three broader measures of association – Pearson’s product-moment correlation, Spearman’s rank correlation, and Kendall’s  $\tau$  – that we will use later to parameterize the relationships that exist within the bulk of a multivariate distribution.

### 2.2.4 Tail Dependence

Tail dependence between a pair of distributions is typically formalized via a conditional tail probability called the coefficients of tail dependence:

**Definition 2.** Let  $X, Y$  be random variables with cdfs  $F_X$  and  $F_Y$ , with  $H$  as their bivariate cdf. The **lower- $u$  tail dependence** of  $H$  is

$$\begin{aligned}\lambda_{L-u} &= P\left(Y < F_Y^{-1}(u) \mid X < F_X^{-1}(u)\right) \\ &= \frac{H(F_X^{-1}(u), F_Y^{-1}(u))}{u}\end{aligned}$$

Note that the bivariate cdf  $H$  is symmetric by its definition, and so lower- $u$  tail dependence is a symmetric property.

Similarly define the **upper- $u$  tail dependence** of  $H$  as

$$\begin{aligned}\lambda_{U-u} &= P\left(Y > F_Y^{-1}(u) \mid X > F_X^{-1}(u)\right) \\ &= \frac{1 - 2u - H(F_X^{-1}(u), F_Y^{-1}(u))}{1 - u}\end{aligned}$$

The limits of these quantities as are the **lower tail dependence** and **upper tail dependence**, respectively (provided that the limits exist):

$$\lim_{u \searrow 0} \lambda_{L-u} = \lambda_L, \quad \lim_{u \nearrow 1} \lambda_{U-u} = \lambda_U.$$

We can use Sklar's theorem to establish tail dependence as a property of the copula, independent of the marginals.

**Theorem 12.** Sklar (1959) Let  $X, Y$  be random variables with cdfs  $F_X$  and  $F_Y$ , with  $H$  as their bivariate cdf. There exists a copula  $C$  such that

$$H(x, y) = C(F_X(x), F_Y(y)).$$

If  $F_X$  and  $F_Y$  are continuous, then  $C$  is unique.

If  $C$  is  $H$ 's corresponding copula in the definition of  $\lambda_{L-u}$  and  $\lambda_{U-u}$ , then by Sklar's Theorem,  $\lambda_{L-u} = \frac{C(u,u)}{u}$  and  $\lambda_{U-u} = \frac{1-2u-C(u,u)}{1-u}$ .

Note that the CDF of the Cube copula,  $F$ , in the bivariate case is given by

$$F(u, v) = \int_0^v \int_0^u f(x, y) dx dy$$

$$= \begin{cases} q_2 uv & \text{if } u \leq a, v \leq a \\ u[q_2 a + q_1(v - a)] & \text{if } u \leq a < v \\ v[q_2 a + q_1(u - a)] & \text{if } v \leq a < u \\ q_2 a^2 + q_1 a[(v - a) + (u - a)] + q_0(u - a)(v - a) & \text{if } u > a, v > a \end{cases}$$

So,

$$\lambda_{L-u} = \begin{cases} q_2 u & \text{if } u \leq a \\ [q_2 a^2 + 2q_1 a(u - a) + q_0(u - a)^2]/u & \text{if } u > a \end{cases}$$

and

$$\lambda_{U-u} = \begin{cases} \frac{1-2+q_2 u}{1-u} & \text{if } u \leq a \\ \frac{1-2u+q_2 a^2+2q_1 a(u-a)+q_0(u-a)^2}{1-u} & \text{if } u > a \end{cases}$$

Note in particular,  $\lambda_L$  and  $\lambda_U$ , the limiting values, are zero. So, the Cube copula allows for precise modeling of  $\lambda_{L-u}$ , but because its density is bounded, the lower tail dependence vanishes at arbitrarily small percentiles. In section 2.4 we propose a method for modeling not just a single  $\lambda_{L-u}$  but a countably infinite set of lower tail dependencies.

### 2.2.5 Measures of Association

Many authors have commented on the shortcomings of Pearson's product-moment correlation,  $\rho_p$ , for measuring associations in copula-based models. The most obvious is that  $\rho_p$  is not defined when marginals have infinite second moments, for example  $t_\eta$  with  $\eta \leq 2$ , and many copula applications use such fat-tailed distributions.

A second shortcoming is that  $\rho_p$  is not a copula property. That is,  $\rho_p$  depends on both the copula and the marginal distributions. This contrasts with Spearman's rank correlation and Kendall's tau,  $\rho_s$  and  $\rho_\tau$  respectively, both of which are copula properties. Here we focus on  $\rho_s$  rather than  $\rho_\tau$ , for reasons described in Sec. 2.3 on copula mixtures.

For a copula with cdf  $F$ , we have  $\rho_s = 12 \int_0^1 \int_0^1 F(x, y) dx dy - 3$ . For the Cube copula then we have

$$\rho_s(F) = 3(a-1)^4 q_0 + 3(a-2)^2 a^2 q_2 - 12a(a-1)^3(a+1)q_1 - 3 \quad (16)$$

Our explicit formula (16) for the Spearman correlation of a Cube copula will prove useful in Sec. 2.3, when we consider mixing the Cube copula with another copula in order to achieve a desired correlation in the bulk of the distribution.

## 2.3 Mixing the Cube Copula

Above we constructed a copula with unusually high tail-dependence; indeed the tail dependence arising from this copula is maximally high within the space

of copulas that have the particularly simple structure we have laid out. However, the correlation with the bulk is zero by construction. How can we incorporate a non-zero bulk correlation structure together with tail dependence? We use a simple result that any mixture of two copulas is again a copula. We can then take a convex combination of a Cube copula with another copula that exhibits bulk correlation, and the resulting copula will exhibit both left tail dependence and bulk correlation.

### 2.3.1 Properties of Copula Mixtures

**Definition 3.** Let  $V$  be a real vector space, and let  $K \subseteq V$  be any convex subset. A function  $f : K \rightarrow \mathbb{R}$  is said to be convex-linear if

$$f(tx + (1-t)y) = tf(x) + (1-t)f(y) \quad \text{for all } t \in [0, 1], x, y \in K.$$

Note that convex-linearity extends naturally to compositions with affine maps in either order. Specifically if  $f : V \rightarrow \mathbb{R}$  is convex-linear and  $\phi : V \rightarrow V$  is an affine map defined by  $\phi(x) = Ax + b$ , then  $f \circ \phi$  is also convex-linear. Similarly if  $\phi' : \mathbb{R} \rightarrow \mathbb{R}$  is affine, then  $\phi' \circ f$  is again convex-linear. These statements remain true when the target space  $\mathbb{R}$  is replaced by an arbitrary vector space, but we will only use the real-valued case.

### 2.3.2 Tail Dependence

Let  $\mathfrak{C}_k$  denote the set of  $k$ -variate copulas; note that  $\mathfrak{C}_k$  is a convex subset of the vector space of all functions from  $\Omega_2 \rightarrow \mathbb{R}$ , hence def. 3 applies. Consider  $\lambda_{L-u}$ , the lower- $u$  tail dependence of the previous section, as a real-valued function defined on  $\mathfrak{C}_2$ .

Let  $C_1$  and  $C_2$  be bivariate copulas, and  $t \in [0, 1]$ . Note that

$$\begin{aligned} \lambda_{L-u}(tC_1 + (1-t)C_2) &= \frac{1}{u} (tC_1 + (1-t)C_2)(u, u) \\ &= t \frac{C_1(u, u)}{u} + (1-t) \frac{C_2(u, u)}{u} \\ &= t\lambda_{L-u}(C_1) + (1-t)\lambda_{L-u}(C_2). \end{aligned}$$

So  $\lambda_{L-u}$  is convex-linear, and the lower-u tail dependence of a mixture copula is the mixture of the component copulas' lower-u tail dependencies. Also convex-linear is  $\lambda_L = \lim_{u \searrow 0} \lambda_{L-u}$ , provided that the limits of the components' exist.

### 2.3.3 Correlations

Let  $(X_1, Y_1)$  and  $(X_2, Y_2)$  be independent continuous random variables with common margins  $F$  (of  $X_1$  and  $X_2$ ) and  $G$  (of  $Y_1$  and  $Y_2$ ). Let  $C_i$  denote the copula of  $(X_i, Y_i)$ . Define

$$Q = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

as in [25]. Then

$$Q = Q(C_1, C_2) = 4 \int C_2 dC_1 - 1 = Q(C_2, C_1). \quad (17)$$

It is immediate from (17) that  $Q$  is convex-linear in each of its two arguments, if the other is held fixed. Let  $\Pi$  denote the independence copula (constant density).

Two popular non-parametric measures of concordance are *Kendall's tau* defined as  $\rho_\tau(C) = Q(C, C)$ , and *Spearman's rho* given by  $\rho_s(C) = 3Q(C, \Pi)$ . From the observation that  $Q$  is convex-linear in each argument, it follows that  $\rho_s(C)$  is convex-linear in  $C$ , while

$$\rho_\tau(tA + (1-t)B)$$

is a polynomial of degree 2 in  $t$ , hence cannot be convex-linear. Similarly, Gini's coefficient

$$\gamma = 2 \int (|u + v - 1| - |u - v|) dC(u, v)$$

and Blomqvist's "medial correlation"  $\beta = 4C(\frac{1}{2}, \frac{1}{2}) - 1$  are also convex-linear functions of  $C$ .

Finally, we remark that the  $n$ -dimensional Spearman's  $\rho_S$ , given by

$$\rho_{S,n} = \frac{n+1}{2^n - (n+1)} \left[ 2^{n-1} \left( \int C d\Pi^n + \int \Pi^n dC \right) - 1 \right]$$

is convex-linear on the convex cone of  $n$ -copulas.

These considerations imply that when dealing with mixture copulas, all of the usual measures of concordance *except Kendall's  $\tau$*  are convex-linear and can be summed across the components of the mixture.

### 2.3.4 The Cube-Gaussian Mixture Copula

Let  $\rho$  be an  $n \times n$  correlation matrix. The *Gaussian copula* with correlation  $\rho$  is defined by its PDF:

$$p_g(\mathbf{u}) = |\rho|^{-1/2} \exp\left[-\frac{1}{2}\zeta'(\rho^{-1} - I)\zeta\right] \quad \text{where} \quad \zeta = \Phi^{-1}(\mathbf{u}).$$

Here  $\Phi$  is the normal CDF applied componentwise to vectors, and  $\mathbf{u} \in [0, 1]^n$ . Building on this we define the *Cube-Gaussian mixture copula* by

$$p_{gt}(\mathbf{u}) = \lambda p_c(\mathbf{u}) + (1 - \lambda)p_g(\mathbf{u}). \tag{18}$$

Here  $\lambda \in [0, 1]$  is the mixture probability. One can view this as a hierarchical model, where a mixing random variable defined on  $\{0, 1\}$  determines which copula that  $X$  will be drawn from.

Eq. (18) is our first example of a copula which has desirable properties for modeling portfolios of risky assets. In the bulk of the distribution, i.e. the region in which assets in the portfolio are not experiencing VaR events individually, the assets behave as though they have correlation matrix  $\rho$ , but if one or more assets is experiencing a VaR event, the conditional probability that others are also seeing their VaR events is much higher than it could be with a Gaussian copula or a t-copula.

One advantage of any copula of the form (18) is that, due to the result in sec. 2.3.3 that many of the standard measures of association are convex-linear on the space of copulas, we see that these measures will be no more difficult to



compute for the Cube-Gaussian copula (18) than for either of its components, and we have already shown how to compute Spearman's correlation for the cube in sec. 2.2.5. The more advanced copulas we will introduce in Sec. 2.4 are also mixtures, and also benefit from the results in sec. 2.3.3.

Eq. (18) also lends itself well to Monte Carlo simulation, since each of the components  $p_c$  and  $p_g$  is easy to simulate. Given a simulation, one can proceed to a full portfolio-level analysis of Value-at-Risk (VaR). The simulation suffices to compute each asset's marginal contribution to portfolio VaR as a numerical derivative. Fig. 6 illustrates the behavior of the Cube-Gaussian mixture via a simulation histogram with  $\rho = 0.3$  bulk correlation,  $a = 0.015$  and maximal tail dependence for these parameters.

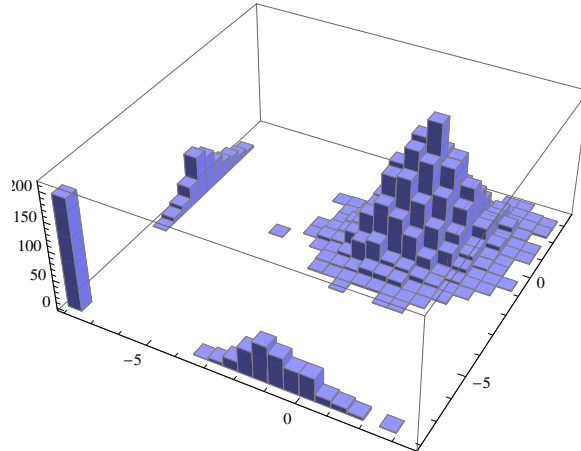


Figure 6: Histogram of a simulation from the Cube-Gaussian mixture with  $\rho = 0.3$  bulk correlation,  $a = (0.015, 0.015)$  and maximal tail dependence.

From Fig. 6 it's intuitively clear that this copula satisfies the two desirable properties laid out in Sec. 2.1. In particular note that there is a large probability density in the simultaneous *lower left* tail, but no corresponding density in the simultaneous *upper right* tail; this is typical of portfolios of financial assets, and completely impossible to achieve with the  $t$ -Copula.

Although surely more realistic for portfolio risk modeling than either the normal copula alone, or the  $t$ -Copula, even Eq. (18) has an important short-coming for the intended application. Fortunately the correction for this short-

coming is known, and leads to interesting further mathematics. This will be the subject of the next section; for the moment, we simply expose the indicated shortcoming.

Note that the probabilities of simultaneous tails beyond the  $a$ -th percentile die off quickly. Suppose for illustration a Cube with  $a = 0.05$ ,  $n = 2$ , and  $q_2 = 16$ . This is equivalent to assuming that the chance of a double tail is twice the chance of a single tail. Then conditional on the first asset having a 95%-VaR event, the probability that both simultaneously have 95%-VaR events (that is,  $\lambda_{L-0.05}$ ) is 0.8. The corresponding probability for a Gaussian copula with  $\rho = 0.9$  is around 0.64. However what if we consider 99%-VaR? Under the Cube,  $\lambda_{L-0.01}$  falls to 0.43, while for the Gaussian is 0.54. Given the Cube's zero asymptotic tail dependence we know this conditional probability converges to zero, but it does so sufficiently quickly to cause concern in some applications, such as in cases where one wants to forecast both 95% and 99% VaRs. The next section introduces a modeling technique, which we call copula nesting, that allows for modeling arbitrarily many points along the tail dependence surface with a sequence of nested Cube copulas.

## 2.4 The Copula Nesting Theorem

In Sec. 2.3, we defined the Cube copula and noted that via selecting  $\mathbf{q}$  and  $a$  appropriately one can precisely specify  $\lambda_{L-a}$ . However, the tail dependence beyond  $a$  degrades, and asymptotes to zero. Suppose that one wants to specify tail dependence at a set of quantile points  $A$ . Figure 7 illustrates the method we propose for doing so. The key observation is that one can nest a second Cube copula within the lower left region of an initial Cube copula. We prove below that the resulting function remains a copula. One can repeat this nesting arbitrarily many times, and in doing so, precisely model  $\lambda_{L-a_n}$  for  $a_n \in A$ .

### 2.4.1 The Nesting Theorem and Proof

Let  $f$  be the Cube copula density on  $\Omega = [0, 1]^n$  with tail parameter  $a$  and values  $q_k$  on the  $k$ -tail for each  $k = 0, \dots, n$ . Let  $s$  be any copula on  $\Omega$ , which we extend to  $\mathbb{R}^n$  by specifying that  $s = 0$  outside  $\Omega$ . Let  $\phi : \Omega \rightarrow \tau_{n,n}$  be

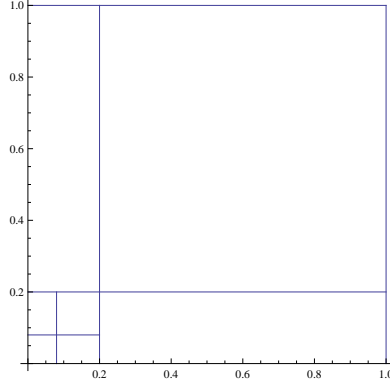


Figure 7: Cell complex underlying a doubly-nested Cube copula in two dimensions. The figure illustrates a cube with  $a = 0.4$  nested inside a cube with  $a = 0.2$ .

the bijective affine map between the indicated regions given by rescaling each coordinate.

Consider the density

$$\hat{s}(\mathbf{u}) = a^{-n} s(\phi^{-1}(\mathbf{u})) \quad (19)$$

Then by our convention  $\hat{s}$  vanishes outside of  $\tau_{n,n}$ . Since we have multiplied by the inverse of the Jacobian, the overall normalization is preserved:  $\int \hat{s}(\mathbf{u}) d\mathbf{u} = 1$ .

In Theorem 13 we construct a copula  $\hat{f}$  which, intuitively, consists of modifying  $f$  by replacing its values in  $\tau_{n,n}$  with a scaled version of  $s$ .

**Theorem 13.** *The multivariate probability density defined by*

$$\hat{f}(\mathbf{x}) = \begin{cases} q_n a^n \hat{s}(\mathbf{x}), & \text{if } \mathbf{x} \in \tau_{n,n} \\ f(\mathbf{x}) & \text{otherwise} \end{cases}$$

with  $\hat{s}$  defined as in Eq. (19), is a copula density.

*Proof.* The scaling is such that the integral over  $\tau_{n,n}$  is unchanged. It follows that  $\hat{f}$  is a probability density. We also claim that  $\hat{f}$  has uniform marginals.

We need to show that the marginal in the  $x_j$  direction is uniform for each  $j = 1, \dots, n$ . For notational simplicity we show this for  $j = n$ ; the same proof holds in each direction. Then we may write  $\mathbf{x} = (\mathbf{y}, x_n)$  where  $\mathbf{y} \in [0, 1]^{n-1}$  and  $x_n \in [0, 1]$ . The marginal function is then

$$m(x_n) = \int_{[0,1]^{n-1}} \hat{f}(\mathbf{y}, x_n) d\mathbf{y}.$$

Note that if  $x_n > a$ , then  $\hat{f}(\mathbf{x}) = f(\mathbf{x})$  and hence  $m(x_n) = 1$  since  $m(x_n)$  is a marginal of the copula  $f$ . Therefore suppose  $x_n \leq a$  and split the integral as follows:

$$\begin{aligned} \int_{[0,1]^{n-1}} \hat{f}(\mathbf{y}, x_n) d\mathbf{y} &= \int_{\tau_{n-1,n-1}} \hat{f}(\mathbf{y}, x_n) d\mathbf{y} + \int_{\tau_{n-1,n-1}^c} \hat{f}(\mathbf{y}, x_n) d\mathbf{y} \quad (20) \\ &= q_n a^n \int_{\tau_{n-1,n-1}} \hat{s}(\mathbf{y}, x_n) d\mathbf{y} + \int_{\tau_{n-1,n-1}^c} f(\mathbf{y}, x_n) d\mathbf{y} \quad (21) \end{aligned}$$

Also in the region  $x_n \leq a$  one has

$$\begin{aligned} q_n a^n \int_{\tau_{n-1,n-1}} \hat{s}(\mathbf{y}, x_n) d\mathbf{y} &= q_n a^n \int_{\tau_{n-1,n-1}} \hat{s}(a\mathbf{z}, x_n) d(a\mathbf{z}) = q_n a^{n-1} \int_{[0,1]^{n-1}} s(\mathbf{z}, x_n/a) d\mathbf{z} \\ &= q_n a^{n-1} = \int_{\tau_{n-1,n-1}} f(\mathbf{y}, x_n) d\mathbf{y} \end{aligned}$$

since  $\int_{[0,1]^{n-1}} s(\mathbf{z}, x_n/a) d\mathbf{z}$  is a marginal of  $s$ .

Plugging this back into the expression (21), we see that the sum (21) collapses into an expression for the marginal of  $f$  at  $x_n$ , which we know to be 1. This completes the proof.  $\square$

*Remark 1.* The same argument also shows that an arbitrary copula can be nested within the 0-tail region  $\tau_{0,n}$ . The Cube copula is essentially the only copula that admits a nesting theorem of this form.

If the nested copula  $s$  is itself a Cube copula, then further copulas can be nested within the inner copula  $s$ . One can in fact do this infinitely-many times, leading to a fractal structure, though for applications in finance or engineering one would typically stop when the tails being modeled are so low-probability that one has no further view on tail dependence or need to model it in those regions.

## 2.4.2 Improving The Cube-Gaussian Mixture

The multiple-nesting property allows the practitioner to customize the copula's tail-dependence properties by specifying as ingredients not only the probability of an  $a$ -quantile VaR event, but also the probabilities of the  $a/10$ -quantile,  $a/100$ -quantile, etc.

In this way we can resolve the fundamental difficulty which plagued the simple form of the Cube-Gaussian mixture discussed in Sec. 2.3.4. With nested copulas, it need not be the case that the conditional probability of an  $n$ -tail  $a$ -quantile event, conditioned on the occurrence of an  $(n - 1)$ -tail  $a$ -quantile event, goes to zero as  $a \rightarrow 0$ . By suitably choosing the  $\mathbf{q}$ -vectors for the inner nested copulas, one can ensure that these probabilities remain bounded away from zero and so that the full copula has a non-zero tail dependence coefficient.

Suppose that in  $n$ -dimensions we have the Cube copula  $p_c(\mathbf{x})$  with parameters  $a, \mathbf{q}$  and we define an *inner Cube copula*  $\hat{p}(\mathbf{x})$  which has the same structure, but different parameters  $\hat{a}, \hat{\mathbf{q}}$  and a normalizing constant set according to eq. (22) below. Note that there are no constraints on  $\hat{a}, \hat{\mathbf{q}}$  aside from the general constraints set by Theorem 11 which apply to all Cube copulas.

As before, we set  $\tau_{k,n}$  to be the  $k$ -tail region of the outer copula. We will use  $\hat{\tau}_{k,n}$  to denote the corresponding regions for the inner copula. Then the nested copula is:

$$p(\mathbf{x}) = \begin{cases} \hat{p}(\mathbf{x}), & \mathbf{x} \in \tau_{n,n} \\ p_c(\mathbf{x}) & \text{otherwise} \end{cases}.$$

The normalizing constant for  $\hat{p}$  is set so that

$$\int_{\hat{\tau}_{n,n}} \hat{p}(\mathbf{x}) d^n \mathbf{x} = q_n a^n. \quad (22)$$

Suppose, as is common in financial risk modeling, we are interested in 95% and 99% VaR, and we wish to build a model with higher-than-normal probabilities of joint tail events occurring at these quantiles. The doubly-nested

Cube achieves this: parameters  $a = 0.05$ ,  $\hat{a} = 0.2$  imply

$$\begin{aligned}\text{Prob}(\text{joint } 95\% \text{ quantile}) &= (0.05)^2 q_n \\ \text{Prob}(\text{joint } 99\% \text{ quantile}) &= (0.01)^2 \hat{q}_n.\end{aligned}$$

This illustrates that the nested copula allows us to tailor the probabilities of these events representing joint observations of extreme outliers. After mixing with the normal, of course the necessary integrals become more difficult to do, but even these can easily be handled numerically.

## 2.5 Literature Review

Sklar introduced the mathematical structure of copulas into the probability and statistics literature in 1959, coining the phrase “copula” with Schweizer in their 1983 textbook [74]. The topic received much attention in decades following its introduction, summarized nicely by Schweizer’s “Thirty Years of Copulas” [73]. This research spawned several introductory papers and textbooks meant to introduce the advanced undergraduate or graduate student to the topic, see for example the appropriately titled “Joy of Copulas” [36], and the excellent texts by Joe [43] and Nelsen [61]. The concept found applications within the fields of engineering and biology, but only recently have researchers applied copulas to economic data. The earliest instances came in the insurance and operations research literature insurance during the mid 1990s. Frees et al [31] in the *Journal of Risk and Insurance* consider the problem of pricing an annuity promised on two lives, and apply Frank’s copula [30], a special case of the Archimedean family of copulas. Jouini and Clemen [44] investigate aggregating expert opinions, also with Frank’s copula. The first mention of copulas within *Management Science* arrives in [94], who study accident “precursors” or “near-misses”, where the joint distribution modeled is that of the failure probability of some safety system under two states of the world depending on whether some other safety system has or has not failed. The first appearance of copulas in an economics journal is also via an investigation into an insurance problem, in the context of a principal agent problem with adverse selection [45]. Overall, the use of copulas in the economics literature has been sparse and very recent. The journals of the American Economic Association record four articles mentioning copulas, all between 2007

and present; *Econometrica* records three mentions, all published in 2010; the *Journal of Political Economy* records one mention [40]; while the *Quarterly Journal of Economics* records none.

The use of copulas in the financial literature was also recent, but has grown explosively in the last ten years. In their widely circulated 1999 working paper, Embrechts, McNeil, and Strauman [26] introduce copulas into modeling financial asset returns. They focus on correcting what they perceive as commonly held views on correlations that “arise from the naive assumption that dependence properties of the elliptical world also hold in the non-elliptical world” and they propose copulas and rank correlations as a remedy. With this background, it is perhaps not surprising that the highly non-elliptical world of credit derivatives emerged as fertile ground for copulas. Li [49] was the earliest published instance, although he cites technical documents from the industry that predate his research (although not explicitly using the copula terminology).<sup>34</sup> Soon after Li’s article, examples of copulas in credit modeling rapidly proliferated; key references are [32] and [72], both of which unify Li’s approach with the latent variable approach of older industry research (KMV and CreditMetrics). Bouye et al [9] provide a reading guide that both introduces the mathematics of copulas and illustrates with applications to credit scoring, asset returns, and risk measurement. Longin and Solnik [53] provide the first published example of copulas used in modeling returns from different equity markets, as well as the first mention of copulas in a top finance journal. They use Gumbel’s copula [39], although interestingly they neither cite the seminal statistical references nor use the phrase “copula” in their paper. Other highly regarded finance journals follow suit: [12] in the *Journal of Financial and Quantitative Analysis*; [5] in the *Journal of Business*; [68] in the *Review of Financial Studies*. Interestingly, the earliest mention of copulas in the *Journal of Financial Economics* is in a footnote to [4], which states that “Embrechts et al. (1999) have recently advocated the use of copulas and rank statistics when measuring dependence in non-normally distributed financial data. However, because the unconditional distributions that we explore . . . are all approximately Gaussian, the linear correlation affords the most natural measure in the present context.” Unfortunately, the evidence presented that the financial data they study are normal concerns only the marginal distribu-

---

<sup>34</sup>Li’s use of the Gaussian copula was pilloried in an article in Wired magazine dubbed “Recipe for Disaster: The Formula That Killed Wall Street”, [71].

tions, and not the joint distribution, which is what is relevant for determining whether to use copulas. Finally, several textbooks provide very thorough introductions to copulas in finance, namely [14] and [84].

One interesting application comes from Rosenberg and Schuermann [69]. They attempt to model the various risks that a complicated financial institution faces (market, credit, and operational) via flexible modeling of marginal distributions each of which is allowed to have a very different shape. The authors conclude that the VaRs of the individual components and the weights that aggregate these components into a portfolio play a larger role in determining portfolio VaR than the choice of the copula. However, the authors consider only Gaussian and Student's-t copulas, which we suspect drives this conclusion.

In general, many of the existing financial applications in the literature seem to view the primary benefit of copulas as simply allowing for arbitrary marginals, without much attention given to the implications of the copula for modeling tail dependence. The emphasis, then, becomes sophisticated modeling of the marginals, with the copula chosen as an afterthought. Rosenberg and Schuermann clearly fits in this category, as do most of the early credit modeling references provided above. A notable exception, and the approach most similar to ours, is Hu's [42], which estimates mixtures of Gaussian, Gumbel, and Gumbel survival copulas using monthly returns from the S&P 500, FTSE 100, Nikkei 225, and Hang Seng. Like our application below, Hu estimates marginals non-parameterically, focusing on the dependence structure rather than the marginals.

## **2.6 An Application to Hedge Fund Returns**

### **2.6.1 Data**

Hedge fund returns provide a natural setting to apply our copula mixture. We think of multivariate hedge fund return distributions as operating under two regimes. In normal times, hedge funds strategies operate with whatever correlations arise naturally from their common exposures to risk factors and correlated trading strategies. However, during stress scenarios, strategies cor-



relate to a much higher degree, as industry-wide balance sheet reductions beget negative returns via market impact, which beget further balance sheet reductions. The causes of these correlated portfolio liquidations can be acute, such as when sudden losses by a large fund become common knowledge (eg Long Term Capital Management in 1998), or more diffuse, such as a contraction in banks' willingness to finance transactions or elevated redemption requests by investors (which are two commonly-cited causes of large hedge fund losses in the fall of 2008). Regardless of cause, the presence of simultaneous deleveragings creates a left tail dependency that can be much more extreme than what one would expect from observing returns during normal times.

To illustrate, we use the Hedge Fund Research indexes (HFRI), which HFR describes as "a series of benchmarks designed to reflect hedge fund industry performance by constructing equally weighted composites of constituent funds, as reported by the hedge fund managers listed within HFR Database." While HFRI returns suffer some serious biases in their construction<sup>35</sup>, they are generally considered the industry standard and have been used in many of the seminal studies of hedge fund returns (Ackerman, McEnally, and Ravenscraft [1]; Liang [50]; Agarwal and Naik [2], [3]; Getmansky, Lo, Makarov [37]; Fung and Hsieh [33]). Specifically, we investigate the joint distribution of the Event Driven (ED) and Relative Value (RV) strategy indexes.<sup>36</sup>

Our data consist of the monthly returns for Event Driven and Relative Value from February 1990 through August 2010, measured in excess of the US 3-month Treasury bill rate. The cumulative returns are shown in Fig. 8, and the scatterplots in Fig. 9. The excess returns are highly correlated, with a  $\rho_S$  of 0.67 over the full sample, despite strategy descriptions that would not suggest such high correlations.

Much of this correlation is due to persistent exposure to common risk factors. We attempted to control for these exposures by OLS regression of each

---

<sup>35</sup>Most seriously, returns are self-reported and funds are free to self-censor.

<sup>36</sup>Event Driven includes as sub-categories: Activist, Credit Arbitrage, Distressed/Restructuring, Merger Arbitrage, Private Issue/Regulation D, Special Situations, and Multistrategy. Relative Value includes as sub-categories: Fixed Income-Asset Backed, Fixed Income-Convertible Arbitrage, Fixed Income-Corporate, Fixed Income-Sovereign, Volatility, Yield Alternatives:Energy Infrastructure, Alternatives:Real Estate, and Multistrategy.

strategy against the excess total returns of several market indexes:

- (a) S&P 500,
- (b) Barclays Capital Aggregate Total Treasury,
- (c) Barclays Capital US Corporate High Yield, and
- (d) S&P Goldman Sachs Commodity.

For each HFR index, all four market indexes (a)–(d) had t-statistics above 2 (with average absolute t-statistics of 6.0 for ED and 4.8 for RV), and adjusted  $R^2$  were 65% for ED and 52% for RV. The Spearman correlation of the residuals to these factors drops to 0.49, but clearly, left-tail correlation remains present even in the residualized data.

### 2.6.2 Methodology

We estimate the parameters of our Cube-Gaussian Mixture copula, described in Section 2.3.4, via a two-step, pseudo-maximum likelihood estimation procedure. First, we estimate marginals via the empirical CDF, and apply an inverse empirical CDF to each variable to transform it into a uniform. In the second step, we estimate the parameters of the copula on these transformed data via maximum likelihood. Alternatively, we could estimate a full information maximum likelihood by specifying marginals, and then maximizing a joint likelihood function that is both a function of the parameters of the marginals and the parameters of the copula. The benefit of using the two-step procedure and non-parametrically estimating marginals is that if parametric marginals are mis-specified and included in a joint likelihood, they will interfere with the copula estimates, which are our focus. Note that the standard errors that arise from this two-step procedure are larger than those that would be naively computed by assuming that the transformed data were the actual data. Intuitively, the inverse empirical CDF does not equal the true inverse CDF, and this source of estimation error must be accounted for in the copula’s parameter estimates. The corrected standard errors come via a “sandwich estimator” of the asymptotic covariance matrix; [35] provide a derivation.

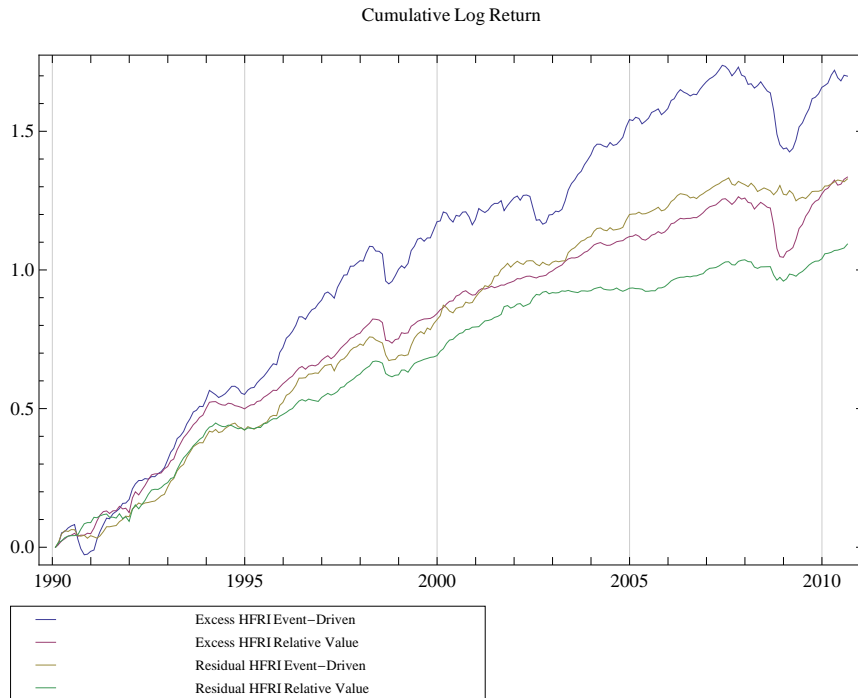


Figure 8: Cumulative Excess Log Returns and Residual Log Returns for HFRI Event Driven and Relative Value Indexes, Monthly from Jan 1990–Aug 2010.

We chose to use a single breakpoint  $a$ , fixed in advance, which defines double-tail region in the Cube, and we estimated the following three parameters:  $q_2$ , which determines the likelihood of the double-tail,<sup>37</sup>  $\lambda$ , the mixing parameter, and  $\rho$ , the correlation of the Gaussian copula. Provided that  $a$  is chosen not to coincide with one of the values of the sample, the likelihood function is differentiable in each parameter, and the maximization was easily solved via Matlab’s `fmincon` function.<sup>38</sup> We estimated with  $a$  equal to 0.01 and 0.05, representing 95% and 99% VaR, two commonly used breakpoints

<sup>37</sup>Note from Section 2.2.2 that  $q_1$  and  $q_0$  are simple functions of  $q_2$ .

<sup>38</sup>In preliminary research, we also experimented with estimating  $a$ , but this proved difficult. The typical solution to this problem involves an  $a$  that coincides exactly with a sample value, which creates a non-differentiability at the maximum likelihood estimate. In fact, not only is the likelihood function not differentiable in  $a$  at such a point, it is not differentiable in  $q_2$  either. This resulted in instability in our numerical routines.

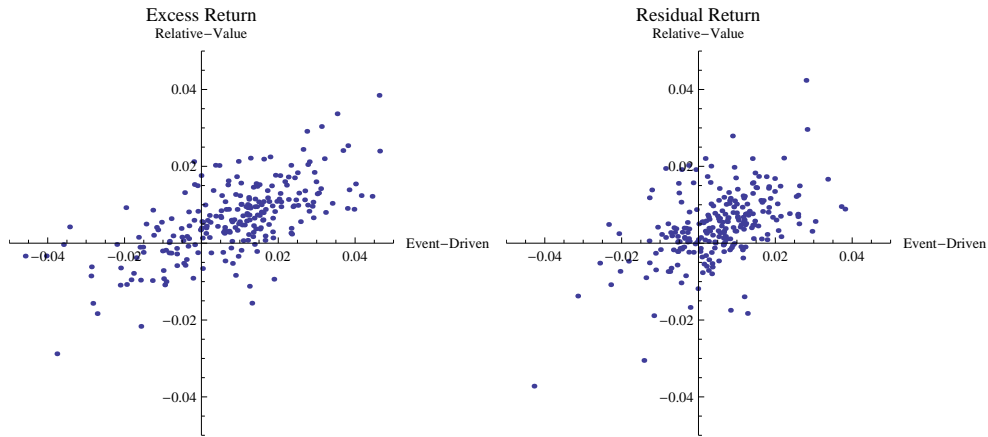


Figure 9: Cumulative Monthly Excess and Residual Log Returns for HFRI Event Driven (horizontal) and Relative Value (vertical) Indexes, Jan 1990-Aug 2010.

in financial risk modeling. Neither 0.01 nor 0.05 was exactly equal to an observation of our sample after it had been mapped to uniform via the inverse empirical CDF, since our sample contained 247 observations.<sup>39</sup>

We also evaluated the fit of the Mixture copula against three competing copulas: pure Gaussian, Student's-t, and Clayton. We estimated the parameters of each of these three copulas via maximum likelihood:  $\rho$  for the Gaussian,  $\nu$  and  $\rho$  for the Student's-t, and  $\theta$  for the Clayton. We evaluate each copula's fit via the value of the log-likelihood and two information criteria that penalize for over-fitting: Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC).

Note that our sample consists of 247 monthly observations. With a sample this size, estimating the 95% and 99% VaRs is imprecise. To see this, note that one can compute a confidence interval for a quantile using the sample order statistics via the binomial distribution. Letting  $X_{(j)}$  denote the  $j$ th largest of a collection of  $n$  iid continuously distributed random variables, and

<sup>39</sup>Had we instead had 100 observations, so that 0.01 and 0.05 would have been elements of our  $U(0,1)$  transformed sample, we could have chosen for example an  $a$  of 0.011 and 0.051 to avoid numerical difficulties. However, choosing 0.011 rather than 0.009 could have a sizable impact on the estimates of the other parameters.

$q_\alpha$  the  $\alpha$ -quantile, we have  $P(X_{(j)} < q_\alpha < X_{(j+1)}) = \binom{n}{j} \alpha^j (1 - \alpha)^{j-1}$ . Starting from the sample  $\alpha$ -quantile, one can expand to successively larger intervals with order statistics as endpoints to obtain confidence intervals with increasing rates of coverage. See Stark [81] for details. In particular, with a sample of 247 observations, a 95% confidence interval for  $q_{0.05}$  centered at the sample 0.05-quantile is contained within  $[x_{(6)}, x_{(19)}]$ , that is, the sample 2.4%-ile to the sample 7.7%-ile. A 95% confidence intervals for  $q_{0.01}$  is contained within  $[-\infty, x_{(5)}]$ , that is, from  $-\infty$  to the sample 2%-ile. If one prefers confidence intervals of finite length, a 90% confidence intervals for  $q_{0.01}$  is contained within  $[x_{(1)}, x_{(6)}]$ ; however there is no interval between the minimum sample observation and maximum sample observation that contains a 95% confidence interval for  $q_{0.01}$ .

However, we don't view this as an impediment to estimating the Cube copula on a dataset of this size. The Cube copula is concerned with the dependence between extreme observations, but not with the specific shape of the tails of a particular marginal distribution. To see this another way, note that our maximum likelihood procedure transforms the data into ranks before the Cube even sees them, so that taking the most extreme observation and scaling it by a factor of 10 along one dimension would have no impact on the Cube's estimates. What is a problem however, is that with a small number of observations, one may not observe extreme dependence. Thus estimating the Cube requires confidence that one has indeed observed extreme dependence within a sample. The uneasiness one should feel making statements about having already observed extreme dependence suggests that in any financial application (and perhaps any application more generally), stress-testing the Cube's parameters (for example for their impact on forecast VaR or contribution to VaR) is critical.

### 2.6.3 Results

First we report the results of the OLS regression of the HFR indexes on the market indexes:

Next we report the estimated parameters of the Cube-Gaussian Mixture, both for excess and residualized HFRI returns:

	HFRI Event Driven	HFRI Relative Value
Intercept	0.005377 (0.000762)	0.004426 (0.000592)
S&P 500	0.1857 (0.0209)	0.0426 (0.0162)
BarCap Agg Total Treasury	-0.1728 (0.0561)	-0.0920 (0.0436)
BarCap US Corp High Yield	0.3172 (0.0332)	0.2549 (0.0258)
S&P GS Commodity	0.0301 (0.0121)	0.04336 (0.0094)
n	247	247
$R^2$	0.655	0.523
Adj- $R^2$	0.650	0.515

Standard errors in parentheses

Note that with  $a = 0.01$ , the corner solution of  $q_2 = a^{-1}$  maximizes the likelihood, and at that estimate the derivative of the likelihood function is undefined, so the approximate standard error is not valid. Given that only 2 observations per variable lie below the 0.01 percentile, we should expect the estimator to be somewhat unstable. The fact that the most extreme observation for HFRI ED residual returns coincides precisely with the most extreme observation for HFRI RV's residual returns <sup>40</sup> means that fitting this highly unlikely (from the perspective of the Gaussian) point perfectly brings a large likelihood gain. This instability would prevent most practitioners from taking the  $a = 0.01$  estimates very seriously, and so we have included them solely to

<sup>40</sup>These occurred during the LTCM crisis of August 1998. The most extreme HFRI ED excess return, also August 1998, occurred with HFRI RV's 2nd most extreme excess return, and vice versa during October 2008.

	$a = 0.05$		$a = 0.01$	
	HFRI Excess Returns	HFRI Residual Returns	HFRI Excess Returns	HFRI Residual Returns
$q_2$	15.45 (37.05)	5.87 (5.01)	100.0 (198.1)	100.0 (149.5)
$\lambda$	0.047 (0.097)	0.393 (0.103)	0.130 (0.090)	0.394 (0.101)
$\rho_s$	0.726 (0.046)	0.751 (0.039)	0.768 (0.037)	0.758 (0.039)
$-\log L$	-82.22	-37.09	-83.07	-38.52

Approximate standard errors in parentheses

illustrate the mechanics of the model.

Finally, we report the values of the likelihood, AIC, and BIC evaluated at the maximum likelihood estimates for the Cube-Gaussian (at  $a = 0.05$  and  $a = 0.01$ ), Gaussian, Clayton, and Student's-t copulas:

On the HFRI Excess returns, Student's-t performs best in terms of the log-likelihood and AIC, but the heavier penalty that BIC enacts on extra estimated parameters causes the pure Gaussian to fit the best of the five models. The Cube-Gaussians both outperform the Clayton on AIC, but are last on BIC due to their extra parameter.

On the HFRI Residual returns, the Cube-Gaussians perform better. BIC still penalizes them sufficiently heavily that they are last, but both are considerably better than the three competing models under AIC's penalty. In terms of intuition for the very different ranking achieved by AIC and BIC, note that AIC's penalty term is a function solely of the number of parameters, namely  $2k$  where  $k$  is the number of estimated parameters, while BIC's penalty term grows with the sample size, via  $k \log(n)$  where  $n$  is the sample size.

Our view of these results is that they at minimum establish that parameter estimation is straightforward for the simplest version of the mixed Cube

	HFRI Excess Returns	HFRI Residual Returns
-log L		
Cube-Gaussian ( $a = 0.05$ )	-82.22	-37.09
Cube-Gaussian ( $a = 0.01$ )	-83.07	-38.52
Gaussian	-82.09	-33.43
Clayton	-78.61	-33.42
Student's-t	-83.92	-36.11
AIC		
Cube-Gaussian ( $a = 0.05$ )	-158.45	-68.19
Cube-Gaussian ( $a = 0.01$ )	-160.13	-71.04
Gaussian	-162.18	-64.87
Clayton	-155.21	-64.83
Student's-t	-163.84	-68.22
BIC		
Cube-Gaussian ( $a = 0.05$ )	-147.92	-57.66
Cube-Gaussian ( $a = 0.01$ )	-149.60	-60.51
Gaussian	-158.67	-61.36
Clayton	-151.70	-61.33
Student's-t	-156.82	-61.20
Cube-Gaussian estimates 3 parameters; Gaussian, 1; Clayton, 1; and Student's-t, 2		

copula, and that it produces results about as good as common alternative approaches to modeling tail dependence. However, we also view the Cube-Gaussian as a modeling tool, rather than just a statistical tool, in the sense that its parameters will generally require both intuition and experience in the arena in which the tool is applied. To make this point obvious, the practitioner who put no thought into the liquidation scenarios that could lead to simultaneous left tails across hedge fund strategies, but instead blindly applied our method, would not have been much better off going into August 1998 than a practitioner who instead used a Gaussian copula.



## 2.7 Conclusions

Motivated by applications to portfolio risk modeling, we searched for a copula which is flexible enough to accommodate a fully-general correlation matrix in the bulk, as well as a very high conditional probability of left-tail events, with no corresponding implication for right-tail events. Our conclusion was that none of the well-known copulas in the literature are quite so flexible. All of these seem to have the property that one can introduce high left tail dependence, but only at the cost of influencing the copula in other ways which make it inadmissible for this sort of risk modeling.

To address this problem, we created a new family of copulas which is more “flexible” in several important ways. It allows one in particular to separately specify the probabilities for  $a$ -quantile events, for a sequence of increasing values of  $a$ , while retaining a fixed correlation structure in the bulk and without requiring the introduction of artificial right-tail dependence.

The new copula we propose takes the form of a nested Cube mixed with a normal copula. The normal copula takes into account the correlation matrix, while the nesting and the cube structure creates a sequence of increasing probabilities of simultaneity in the left tails. The parameter vector  $\mathbf{q}$  can be tuned to make these conditional probabilities as large as mathematically possible. Since the  $q_i$ 's are subject only to simple linear constraints, it is not difficult to tune the  $\mathbf{q}$ -vector while maintaining consistency.

Theorem 11 and Theorem 13 ensure that the resulting mixtures are indeed copulas. This structure has several desirable properties, including the fact that certain important and widely-used measures of concordance, such as Spearman's rho, distribute over mixtures and can be computed explicitly for the components.

Applications in financial engineering include a better estimation of VaR and contributions to VaR for portfolios of assets which have moderate correlation in normal times, but which tend to experience highly correlated drawdowns in crisis periods.

## References

- [1] 1999. Ackermann, C.; McEnally, R.; Ravenscraft, D.: *The Performance of Hedge Funds: Risk, Return, and Incentives*. Journal of Finance, Volume 54, Issue 3, Jun, pp. 833-874.
- [2] 2000. Agarwal, C.; Naik, N.: *Multi-Period Performance Persistence Analysis of Hedge Funds*. Journal of Financial and Quantitative Analysis, Volume 35, Issue 3, pp. 327-342.
- [3] 2000. Agarwal, C.; Naik, N.: *Generalized Style Analysis of Hedge Funds*. Journal of Asset Management, Volume 1, Issue 1, Jul, pp. 93-109.
- [4] 2001. Andersen, T.; Bollerslev, T.; Diebold, F.; Ebens, H.: *The Distribution of Realized Stock Return Volatility*. Journal of Financial Economics, Volume 61, pp 43-76.
- [5] 2003. Ane, T.; Kharoubi, C.: *Dependence Structure and Risk Measure*. Journal of Business, Volume 76, pp 411-438.
- [6] 2002. Ariely, D.; Wertenbroch, K.: *Procrastination, Deadlines, and Performance: Self-Control by Precommitment*. Psychological Science, Volume 13, pp 219-224.
- [7] 1997. Baddeley, A.: *Human Memory: Theory and Practice*. Psychology Press.
- [8] 2003. Bernheim, D.; Thomadsen, R.: *Memory and Anticipation*. Mimeo, Stanford University.
- [9] 2000. Bouye, E.; Durrelman, V.; Nikeghbali, A.; Riboulet, G.; Roncalli, T.: *Copulas for Finance: A Reading Guide and Some Applications*. Unpublished working paper.
- [10] 2002. Brewer, N.; Keast, A.; Rishworth, A.: *Improving the Confidence-Accuracy Relation in Eyewitness Identification: Evidence from Correlation and Calibration*. Journal of Experimental Psychology: Applied, Volume 8, pp. 44-56.

- [11] 1989. Camerer, C.; Loewenstein, G.; Weber, M.: *The Curse of Knowledge in Economic Settings: An Experimental Analysis*. Journal of Political Economy, Volume 97, pp. 1232-1254.
- [12] 2003. Chen, R.; Sopranzetti, B.: *The Valuation of Default-Triggered Credit Derivatives*. Journal of Financial and Quantitative Analysis, Volume 38, pp 359-382.
- [13] 2005. Chen, Y.; Moorthy, S.; Zhang, J.: *Price Discrimination After The Purchase: A Note on Rebates as State-Dependent Discounts*. Management Science, Volume 51, pp. 1131-1140.
- [14] 2004. Cherubini, U.; Luciano, E.; Vecchiato, W.: *Copula Methods in Finance*, Wiley.
- [15] 2001. Choi, J.; Laibson, D.; Brigitte, M.; Metrick, A.: *Saving for Retirement on the Path of Least Resistance*. In: *Behavioral Public Finance: Toward a New Agenda*, ed. McCaffrey, E.; Slemrod, J.: Russell Sage Foundation.
- [16] 2003. Chuang, T.: *Rebates: The discount that makes people see red*. Orange County Register, November 17.
- [17] 1988. Cockburn, J.; Smith, P.: *Effects of Age and Intelligence on Everyday Memory Tasks*. In: *Practical Aspects of Memory: Current Research and Issues, Volume 1: Memory in Everyday Life*, ed: Gruneberg, M.; Morris, P.; Sykes, R.: John Wiley & Sons.
- [18] 1993. Cohen, G.: *Memory and Ageing*. In: *Memory in Everyday Life*, ed. Davies, G.; Logie, R.: North-Holland.
- [19] 1996. Cook, P.; Frank, R.: *The Winner-Take-All Society*. Penguin Books.
- [20] 2006. DellaVigna, S.; Malmendier, U.: *Paying Not to Go to the Gym*. American Economic Review, Volume 96, pp. 694-719.
- [21] 1885. Ebbinghaus, H.: *ber das Gedchtnis*. In: *Untersuchungen zur experimentellen Psychologie*, Duncker & Humblot.
- [22] 2009. Edwards, M.: *The Rebate "Rip-Off": New York's Legislative Responses to Common Consumer Rebate Complaints*. Pace Law Review, Volume 29, pp. 471-478.

- [23] 1988. Ellis, J.: *Memory for Future Intentions: Investigating Pulses and Steps*. In: *Practical Aspects of Memory: Current Research and Issues, Volume 1: Memory in Everyday Life*, ed: Gruneberg, M.; Morris, P.; Sykes, R.: John Wiley & Sons.
- [24] 1993. Ellis, J.; Nimmo-Smith, I.: *Recollecting Naturally-Occurring Intentions: A Study of Cognitive and Affective Factors*. *Memory*, Volume 1, 107-126.
- [25] 2003. Embrechts, P; Lindskog, F; McNeil, A.: *Modelling Dependence with Copulas and Applications to Risk Management*. In: *Handbook of Heavy Tailed Distributions in Finance*, ed. S. Rachev, Elsevier, Chapter 8, pp. 329-384.
- [26] 1999. Embrechts, P; McNeil, A.; Straumann, D.: *Correlation and Dependence in Risk Management: Properties and Pitfalls*. Unpublished working paper.
- [27] 2010. Ericson, K.: *Forgetting We Forget: Overconfidence and Prospective Memory*. *Journal of the European Economic Association*, forthcoming.
- [28] 2010. Ericson, K.: *On Memory and Procrastination*. Mimeo, Harvard University.
- [29] 1975. Fischhoff, B.: *Hindsight / Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty*. *Journal of Experimental Psychology: Human Perception and Performance*, Volume 1, pp. 288-299.
- [30] 1979. Frank, M.: *On the simultaneous associativity of  $F(x, y)$  and  $x + y - F(x, y)$* . *Aequationes Mathematicae*, Volume 19, 194-226.
- [31] 1996. Frees, E.; Carriere, J.; Valdez, E.: *Annuity Valuation with Dependent Mortality*. *Journal of Risk and Insurance*, Volume 63, pp. 229-261.
- [32] 2001. Frey, R.; McNeil, A.; Nyfeler, M.: *Copulas and Credit Models*. *RISK*, October, pp 111-114.
- [33] 1999. Fung, W.; Hsieh, D.: *A Primer on Hedge Funds*. *Journal of Empirical Finance*, Volume 6, pp. 309-331.
- [34] 1976. Gates, S.; Colburn, D.: *Lowering Appointment Failures in a Neighborhood Health Center*. *Medical Care*, Volume 14, pp. 263-267.

- [35] 1995. Genest, C.; Ghoudi, K.; Rivest, L.-P.: *A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions*. *Biometrika*, Volume 82, pp 543-552.
- [36] 1986. Genest, L.; MacKay, J.: *The Joy of Copulas: Bivariate Distributions with Uniform Marginals*. *American Statistician*, Volume 40, pp. 280-283.
- [37] 2004. Getmansky, L.; Lo, A.; Makarov, I.: *An Econometric Analysis of Serial Correlation and Illiquidity in Hedge-Fund Returns*. *Journal of Financial Economics*, Volume 74, pp. 529-609.
- [38] 2005. Grow, B.; Chatwal, R.: *The Great Rebate Runaround*. *Business Week*, December 5.
- [39] 1961. Gumbel, E.: *Bivariate Logistic Distributions*. *Journal of the American Statistical Association*, Volume 56, pp 335-349.
- [40] 2003. Haile, P.; Tamer, E.: *Inference with an Incomplete Model of English Auctions*. *Journal of Political Economy*, Volume 111, pp. 1-51
- [41] 1984. Harris, J.: *Remembering to do things: A forgotten topic*. In: *Everyday Memory, Actions and Absent-Mindedness*, ed. Jeff E. Harris and Peter E. Morris, Academic Press.
- [42] 2006. Hu, L.: *Dependence Patterns Across Financial Markets: A Mixed Copula Approach*. *Applied Financial Economics*, Volume 16, pp 717-729.
- [43] 1997. Joe, H.: *Multivariate Models and Dependence Concepts*, CRC Press.
- [44] 1996. Jouini, N.; Clemen, R.: *Copula Models for Aggregating Expert Opinions*. *Operations Research*, Volume 44, pp. 444-457.
- [45] 1999. Landsberger, M.; Meilijson, I.: *A General Model of Insurance under Adverse Selection*. *Economic Theory*, Volume 14, pp. 331-352.
- [46] 1990. Leonesio, R.; Nelson, T.: *Do Different Metamemory Judgments Tap the Same Underlying Aspects of Memory?*. *Journal of Experimental Psychology: Learning, Memory & Cognition*; Volume 16, pp. 464-470.

- [47] 1977. Levy, R.; Claravall, V.: *Differential Effects of a Phone Reminder on Appointment Keeping for Patients with Long and Short Between-Visit Intervals*. *Medical Care*, Volume 15, pp. 435-438.
- [48] 1984. Levy, R.; Loftus, G.: *Compliance and Memory*. In: *Everyday Memory, Actions and Absent-Mindedness*, ed. Jeff E. Harris and Peter E. Morris, Academic Press.
- [49] 2000. Li, D.: *On Default Correlation: A Copula Function Approach*. *Journal of Fixed Income*, Volume 9, pp 43-54.
- [50] 1999. Liang, B.: *On the Performance of Hedge Funds*. *Financial Analysts Journal*, Volume 55, Issue 4, Jul-Aug, pp. 72-85.
- [51] 2003. Loewenstein, G.; O'Donoghue, T.; Rabin, M.: *Projection Bias in Predicting Future Utility*. *Quarterly Journal of Economics*, Volume 118, pp. 1209-1248.
- [52] 1979. Loftus, E. F. *Eyewitness Testimony*. Harvard University Press.
- [53] 2001. Longin, F.; Solnik, B.: *Extreme Correlation of International Equity Markets*. *Journal of Finance*, Volume 56, pp 649-676.
- [54] 2001. Madrian, B.; Shea, D.: *The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior*. *Quarterly Journal of Economics*, Volume 116, 1149-1187.
- [55] 2002. McLaughlin, K.: *Claiming That Holiday Rebate: Is it Really Worth the Headache?*. *Wall Street Journal Online*, December 3.
- [56] 2007. McDaniel, M.; Einstein, G.: *Prospective Memory: An Overview and Synthesis of an Emerging Field*. Sage Publications.
- [57] 1982. Meacham, J.; Leiman, B.: *Remembering to Perform Future Actions*. In: *Memory Observed: Remembering in Natural Contexts*, ed. Harris, J.; Morris, P.: H. Freeman and Company.
- [58] 1977. Meacham, J.; Singer, J.: *Incentive Effects in Prospective Remembering*. *The Journal of Psychology*, Volume 97, pp. 191-197.

- [59] 1984. Morris, P.: *The Validity of Subjective Reports on Memory*. In: *Everyday Memory, Actions and Absent-Mindedness*, ed. Jeff E. Harris and Peter E. Morris, Academic Press.
- [60] 2002. Mullainathan, S.: *A Memory-Based Model of Bounded Rationality*. *Quarterly Journal of Economics*, Volume 117, pp. 735-774.
- [61] 1998. Nelsen, R.: *An Introduction to Copulas*, Springer.
- [62] 1999. O'Donoghue, T.; Rabin, M.: *Incentives for Procrastinators*. *Quarterly Journal of Economics*, Volume 114, pp. 769-816.
- [63] 1999. O'Donoghue, T.; Rabin, M.: *Doing it Now or Later*. *American Economic Review*, Volume 89, pp.103-124.
- [64] 1970. Orne, M.: *Hypnosis, Motivation, and the Ecological Validity of the Psychological Experiment* In: *Nebraska Symposium on Motivation Vol.*, ed. Arnold, W.; Page, M.: University of Nebraska Press
- [65] 2008. Palmer, K.: *Why Shoppers Love to Hate Rebates*. *US News and World Report*, January 18.
- [66] 1959. Parker, J.; Taylor, E.; Barnett, R.; Martens, S.: *Rating Scale Content: The Relationships Between Supervisory- and Self-Rating*. *Personnel Psychology*, Volume 12, pp. 49-63.
- [67] 1997. Piccione, M.; Rubinstein, A.: *On the Interpretation of Decision Problems with Imperfect Recall*. *Games and Economic Behavior*, Volume 20, pp. 3-24.
- [68] 2004. Poon, S.; Rockinger, M.; Tawn, J.: *Extreme Value Dependence in Financial Markets: Diagnostics, Models, and Financial Implications*. *Review of Financial Studies*, Volume 17, pp 581-610.
- [69] 2006. Rosenberg, J.; Schuermann, T.: *A General Approach to Integrated Risk Management with Skewed, Fat-Tailed Risks*. *Journal of Financial Economics*, Volume 79, pp 569-614.
- [70] 1996. Rubin, D.; Wenzel, A.: *One hundred years of forgetting: A quantitative description of retention*. *Psychological Review*, Volume 103, pp. 734-760.

- [71] 2009. Salmon, F.: *Recipe for Disaster: The Formula that Killed Wall Street*. Wired Magazine, February 23.
- [72] 2001. Schonbucher, P; Schubert, D.: *Copula-Dependent Default Risk in Intensity Models*. Working paper, Department of Statistics, Bonn University.
- [73] 1991. Schweizer, B.: *Thirty Years of Copulas*, in ed: Dall’Aglia, G.; Kotz, S.; Salinetti, G.: *Advances in Probability Distributions with Given Marginals: Beyond the Copulas*, Springer.
- [74] 1983. Schweizer, B.; Sklar, A.: *Probabilistic Metric Spaces*. North-Holland, Amsterdam.
- [75] 1992. Shafir, E.; Tversky, A.: *Choice Under Conflict: The Dynamics of Deferred Decision*. Psychological Science, Volume 3, pp. 358-361.
- [76] 1976. Shepard, D.; Moseley, T.: *Mailed Versus Telephoned Appointment Reminders to Reduce Broken Appointments in a Hospital Outpatient Department*. Medical Care Volume 14, pp. 268-273.
- [77] 2004. Silk, T.: *Examining Purchase and Non-Redemption of Mail-In Rebates: The Impact of Offer Variables on Consumers’ Subjective and Objective Probability of Redeeming*. Unpublished dissertation, University of Florida.
- [78] 2003. Silk, T.; Janiszewski, C.: *Managing Mail-In Rebate Promotions*. Mimeo, University of South Carolina.
- [79] 1959. Sklar, A.: *Fonctions de repartition a n dimensions et leurs marges*. Publications de l’Institut de Statistique de l’Universite de Paris, Volume 8, pp. 229-231.
- [80] 2002. Spencer, J.: *Rejected! Rebates Get Harder To Collect*. Wall Street Journal, June 11.
- [81] 2010. Stark, P.; *SticiGui*, <http://statistics.berkeley.edu/stark/SticiGui/index.html>.
- [82] 1981. Svenson, O.: *Are We All Less Risky and More Skillful Than Our Fellow Drivers?*. Acta Psychology, Volume 47, pp. 143-148.



- [83] 2000. Thaler, R.: *From Homo Economicus to Homo Sapiens*. Journal of Economic Perspectives, Volume 14, pp. 133-141.
- [84] 2007. Trivedi, P.; Zimmer, D.: *Copula Modeling: An Introduction for Practitioners*, Now Publishers.
- [85] 2006. Tugend, A.: *A Growing Anger Over Unpaid Rebates*. New York Times, March 4.
- [86] 1985. Vesonder, G.; Voss, J.: *On the Ability to Predict One's Own Responses While Learning*. Journal of Memory and Language, Volume 24, pp. 363-376.
- [87] 2004. Whitaker, B.: *A Web Offer Too Good To Be True? Read the Fine Print*. New York Times, December 26.
- [88] 1976. Wilkins, A.: *A Failure to Demonstrate the Effects of the "Retention Interval" in Prospective Memory*. Mimeo.
- [89] 1978. Wilkins, A.; Baddeley, A.: *Remembering to Recall in Everyday Life: An Approach to Absentmindedness*. In *Practical Aspects of Memory*, ed. Gruneberg, M.; Morris, P.; Sykes, R.: Academic Press.
- [90] 2009. Willis, G.: *Free trial or Internet scam?*. CNNMoney.com, December 17.
- [91] 2003. Wilson, A.: *Bounded Memory and Biases in Information Processing*. Mimeo, Princeton University
- [92] 1989. Wilson, B.; Cockburn, J.; Baddeley, A.; Hiorns, R.: *The Development and Validation of a Test Battery for Detecting and Monitoring Everyday Memory Problems*. Journal of Clinical and Experimental Neuropsychology Volume 11, pp 855-870.
- [93] 1979. Wylie, R.: *The Self-Concept (Vol.)* The University of Nebraska Press.
- [94] 1998. Yi, W.; Bier, V.: *An Application of Copulas to Accident Precursor Analysis*. Management Science, Volume 44, pp. S257-S270.

- [95] 2004. Zaidi, F.: *Exploiting Consumer Inattention in the Marketplace: Camouflage Pricing with Negative-Option Subscriptions*. Mimeo, University of California-Berkeley.