Evaluating Experience-Rated Incentive Schemes

by

John Treble
(Swansea University)
August 2009

Evaluating Experience-Rated Incentive Schemes

by

John Treble[1]

This paper introduces what I believe to be a new technique for the analysis of a particular class of incentive schemes. The schemes I have in mind are systems where the constraints confronting a decision-maker vary from time to time according to a record of their past behaviour. These are often referred to as 'experience rated' schemes, and they are widely used, not only in labour markets, but also in insurance, legal systems, driver licensing, sports leagues and retail. Experience rating is, for instance, one of the main tools used by firms to create switching costs and encourage brand loyalty[2]. The discussion in the present paper is restricted to a subset of experience-rated schemes. I call this the subset of *redemptive* schemes, because they have the property that whatever state of grace or disgrace the decision-maker may currently be in, it is always possible to exit it, and return to it given an appropriate pattern of behaviour.

1. Redemptive and non-redemptive experience-rating

The UK driving licence points scheme specifies that drivers who violate traffic laws can have points deducted from their licence, up to a maximum of 10. Once more than 10 points have been deducted, the licence is withdrawn for a prescribed period. Many kinds of violation are possible, and the number of points deducted varies

[1] I have discussed the model in this paper on many occasions with Tim Barmby. Martin Dufwenberg suggested the connection with time consistency. Tony Lancaster pointed me towards the literature on serial correlation. Mike Nolan and Assad Jalali have been effectively dubious. The verandah of The Big Blue House in Tucson, Arizona is a very congenial place for doing algebra, even in July. Any remaining errors are, of course, my responsibility.

[2] Klemperer(1995). Unfortunately, Fudenberg and Tirole (2000) have introduced the alternative phrase 'behavior-based' to mean the same thing as experience-rated.

according to the nature of the violation. Points deductions are not permanent. The points are restored after three years have elapsed. The scheme is thus redemptive.

Many US states now have an habitual offender law, commonly known as 'three strikes and you're out'. Under these laws, those who are convicted of three felonies, are imprisoned for life automatically[3]. As far as I know, there is no limitation on the longevity of relevant convictions. The scheme is thus non-redemptive.

A third example, which is worked out in full below, concerns a snack-food packaging plant in Ashby-de-la-Zouch. Here, workers who take more than one spell of absence in a rolling 5-week period, are banned from eligibility for overtime for a period of 4 weeks following the end of the most recent spell. This is similar to the driving licence scheme in that 2 spells trigger a penalty, but spells are not counted once they are more than 5 weeks in the past. This is a redemptive scheme.

Such schemes can be represented in mathematical form. Suppose drivers make a daily decision as to the recklessness of their driving, and that monitoring is imperfect. They can then be seen as controlling the probability that they are caught speeding. We may also suppose that their choice is conditioned by the consequences of a detected violation, so that a driver with 3 speeding convictions in the 3 years prior to date t (who could be fined £60 and whose licence can be withdrawn for a prescribed period), may make a different decision than that made by a driver with none (who could be fined £60 and have 3 points recorded on their licence). Furthermore, the decision may be conditioned by the length of time until the expiry of existing points, so that a driver with 3 points expiring tomorrow, may be more willing to risk being detected speeding than one with 3 points expiring in 2 years' time.

These elements of the decision problem can be used to define a set of *states*. In the driving licence example, they may be indexed by i) whether or not a driver has a licence; ii) in the event that the driver has a licence, the current number of points deducted ($n = \{0, 3, 6, 9\}$), and the dates of their deduction; iii) in the event that the

[3]. LA Times, October 28, 1995, p1: A homeless parolee convicted for breaking into a Santa Ana restaurant and stealing four cookies was sentenced to life in prison Friday after the judge said she had no choice under the state's "three strikes" law. Orange County Superior Court Judge Jean Rheinheimer acknowledged the three-strikes law was "a harsh one" but left her no alternative. Weber had previously been convicted of burglary and assault with a deadly weapon.

driver has had the licence withdrawn, the date of restoration. As time passes, the driver moves from state to state, with certain probabilities. Thus, a licensed driver who is in state 'zero points', could stay in this state, or with some probability transit to '3 points with 3 years to run'[4]. The following day, this driver could transit either to '3 points with 2 years, 364 days to run' or to '6 points, 3 with 3 years, and 3 with 2 years, 364 days to run'.

There are, of course, many models that could be used to represent this structure: the degree of time aggregation is one key choice, since disaggregation can increase the number of states greatly. For practical purposes, using a daily model to analyse driving behaviour is probably neither sensible nor feasible. The main point here, though, is that the rules of the experience-rating can be used to construct a Markov matrix, describing the probabilities of transition from state to state.

The three-strikes law can be similarly modelled. The resulting Markov matrix is much smaller, since there are only 4 possible states: strikes against $= \{0,1,2,3\}$.

The relative size of the matrices generated by these two schemes is unimportant in the present context. There is, however, a second difference between them. The Markov matrix representing the three-strikes law contains an absorbing state (3 strikes against), the other does not. This is the formal difference between redemptive and non-redemptive experience-rating.

The rest of the paper assumes redemptive experience-rating.

2. Likelihoods for data from experience-rated systems

Experience-rating induces a dynamic programming problem that I assume agents solve to determine their optimal responses. The conventional approach to such problems is to suppose that each agent has a utility function and a discount rate. They look forward, assess the response of their environment, and their own responses to that response, and make a decision based on this knowledge. Thus, a car-driver knowing their own propensity for lack of care in speed control may drive more carefully after just three points have been lost, since they know that they have only 7 points left to lose, where before they had 10. Care taken might increase if three, or six, more points are lost, and decrease again when the first three lost points are restored.

[4] Two or more violations on the same day are possible, too, but for simplicity's sake we ignore this, as well as the complications introduced by leap years.

These considerations suggest that the transition probabilities are, at least partially, within the control of the agent. I write the Markov matrix with the probabilities in each row controlled by the agent by means of a single parameter, $\tilde{\sigma}_i$. Iid random draws are taken each period from a distribution, $\Phi(\sigma)$, with associated density, $\phi(\sigma)$. For example, in the absenteeism model analysed below, workers choose optimal reservation levels of morbidity which become higher as the threat of withdrawal of eligibility for overtime increases.

The Markov matrix for such a problem, where each decision is binary, is a stochastic matrix $\boldsymbol{C} = \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_K \end{bmatrix}$ where C_i is a vector with positive elements in rows *r* and *s* only, such that $c_{is}(\tilde{\sigma}_i) = 1 - c_{ir}(\tilde{\sigma}_i)$. This matrix not only encodes the rules of the experience-rating, but with appropriate choice of the vector $\{\tilde{\sigma}_i\}_{i=1}^K$, it can also be used to describe the optimal behaviour of an agent making decisions within the scheme.

Defining $c_{0i}(\tilde{\sigma}_i)$ as the contemporaneous utility received during a one-period occupancy of state *i*, and assuming that the decision-maker discounts future utility exponentially, the optimisation problem can now be stated as:

$$\text{Choose } \{\tilde{\sigma}_i\}_{i=1}^K \text{ to maximise } EV_i = c_{0i}(\tilde{\sigma}_i) + \sum_{k=1}^K \delta c_{ik}(\tilde{\sigma}_i) EV_k, \; i = 1,...K \tag{1}$$

The solution to this problem is described in Proposition I.

<u>Proposition I:</u> Let $\boldsymbol{A} = \boldsymbol{I} - \delta\boldsymbol{C}$, $\boldsymbol{B} = \boldsymbol{A}^{-1}$ with generic element b_{ij}. Let $\boldsymbol{X}_{\bullet n=y}$ denote the matrix $\boldsymbol{X}$ with its *n*'th column replaced by the vector *y*, and $\boldsymbol{X}_{n\bullet=y}$ denote the matrix $\boldsymbol{X}$ with its *n*'th row replaced by the vector *y*. Define $\boldsymbol{\Psi} = \left\{ \frac{\partial c_{ik}}{\partial \tilde{\sigma}_i} \right\}_{i,k=1}^K$. Then for each *n* the optimal vector of reservation levels $\{\tilde{\sigma}_i\}_{i=1}^K$ is the solution to:

$$\left\{U\left(0,\tilde{\sigma}_i\right)-U\left(1,\tilde{\sigma}_i\right)\right\}_{i=1}^{K}=\delta\left\{\frac{1}{b_{ni}\phi\left(\tilde{\sigma}_i\right)}\right\}_{i=1}^{K}\odot\left|A_{\bullet n=c_0}\right| diag\left[\left\{\left(\left[\left(A_{\bullet n=c_0}\right)^{-1}\right]'\right)_{\bullet n=0}-\left(A^{-1}\right)'\right\}\Psi'\right] \quad (2)$$

Proof: See Appendix I.

Equation (2) looks formidable, but it has some nice properties:

i) *Time consistency*. Strotz(1956) argues that exponential discounting is the unique discounting régime ensuring intertemporal consistency in decision-making, and so it appears here. Indeed, Strotz's theorem could probably be seen as constituting a proof of Proposition 2. Nonetheless, it seems worthwhile to check by calculation, mainly because the states in our problem are revisited at random times.

Proposition II: For each *i*, $U\left(0,\tilde{\sigma}_i\right)-U\left(1,\tilde{\sigma}_i\right)$ is independent of the choice of *n*, when exponential discounting is used.

Proof: See Appendix II.

ii) *Generalisation to non-exponential discounting.* I do not know the full range of non-exponential discounting schemes that have been discussed in the literature. The leading alternative in current literature is the quasi-hyperbolic scheme introduced by Laibson(1997), which has the virtue of parsimoniousness, since it adds only one parameter over and above those involved in a model with exponential discounting. The model described above can be extended in a straightforward manner to take account of this generalisation. (See Appendix III) Presumably other non-exponential schemes can also be accommodated.

iii) *Construction of the likelihood.* The solutions to equations (2) (or their non-exponential equivalents) specify the optimal policy to be followed during occupancy of each state. They can therefore generally be used to construct a likelihood for data representing the behaviour of individual decision-makers over time. The details of how this may be done probably vary with the particular applications being studied. An example follows.

3. Example: Absence Control in rural Leicestershire

Models of the kind described in this paper require the use of two types of data for effective estimation. The first is the familiar type of data set consisting of a record of decision-makers' behaviour under the experience-rated scheme. The second is the rules of the scheme. These two types of data are handled in different ways. The rules of the scheme are incorporated into the structure of the likelihood function, which can then be maximised using the record of behaviour.

To make clear how this works, I present an example worked out in full. An employer in Leicestershire operates an experience-rated absence control system. Specifically, workers who take more than two spells of absence in a rolling five week period, will be ineligible for overtime during the four weeks following the end of the last spell.

I now show how these rules can be used to derive the transition matrix ***C*** for this problem, assuming that daily data are available in the record of behaviour. In this framework, ***C*** is of order 86.[5]

To begin, I construct a dynamic binary choice model of absence behaviour.

Suppose utility takes the form $U = U\left(a, b \mid \theta\right)$ where a is zero if the agent makes the first of two available choices and one if s/he makes the second; b is the current realisation of a random shock, σ and θ is a vector of parameters. Under reasonably general conditions a reservation level of σ, $\tilde{\sigma}$, exists such that $U(0,\tilde{\sigma} \mid \theta) = U(1,\tilde{\sigma} \mid \theta)$. The agent's decision rule is then: '$a = 0$ if $\sigma_t < \tilde{\sigma}$; $a = 1$ if $\sigma_t > \tilde{\sigma}$'.

The $\tilde{\sigma}$ defined above is appropriate where there are no future consequences of today's decision. When experience-rating is present, however, there will be a future penalty with expected value $\Pi > 0$. In this case the reservation value of σ must satisfy

$$U(0,\tilde{\sigma} \mid \theta) = U(1,\tilde{\sigma} \mid \theta) - \Pi \tag{3}$$

[5] I ignore weekend days, which are not worked in the plant from which our data are drawn. Strictly speaking, when I do this I should apply treble the discount rate to Mondays, but we have not got around to doing this yet, and I doubt if it makes much difference.

which implies $U(0,\tilde{\sigma} \mid \theta) < U(1,\tilde{\sigma} \mid \theta)$. It is simple to show that $\tilde{\sigma}$ is a monotonic increasing function of Π, so that the larger the future penalty, the greater $\tilde{\sigma}$ will be at the margin between attendance and absence.

The reservation level is important in modelling behaviour because it determines, *ceteris paribus*, the probability that the agent makes choice 1, and can therefore be used to construct the likelihood function. If σ has probability distribution $\Phi(\sigma)$, the probability of an absence is $1-\Phi(\sigma)$. Computation of Π's and σ's is conceptually straightforward, but complex in practice. It is done by taking the derivatives of the relevant value function with respect to the vector of $\tilde{\sigma}$'s, setting them equal to zero, and solving. We now turn to that problem.

The expected value of the optimal policy to an agent who is in state *j* can be written in the form of Equation (1)

$$EV_j = c_{0j} + \sum_{k=1}^{K} \delta c_{jk} EV_k, \quad j = 1,...,K. \tag{4}$$

Here, c_{0j} is the contemporaneous utility of the optimal policy during the current period for a worker in state j, while the summation is the discounted expected utility during the remainder of time. For $j,k = 1,...,K$, c_{jk} denotes the probability of transition from state j to state k. Let c_0 denote the vector $\{c_{0j} \mid j = 1,...,K\}$ and $\boldsymbol{C}$ denote the transition matrix $\{c_{jk} \mid j,k = 1,...,K\}$. The parameter δ is a discount factor. The problem is solved for this version, which uses exponential discounting, using Proposition I, and can be reformulated for agents who use quasihyperbolic discounting and solved using Proposition III.

To construct ***C***, we use the rules of the scheme to determine a state space for the problem. States are distinguished by the number of absence spells a worker has had in the last 5 weeks (0, 1 or >1), by the number days that have elapsed since the most recent spell ended, and by behaviour on the previous day. Imagine a worker moving through time. He can either have no absence spells in the last five weeks (*F* for Free), one absence spell (*S* for one Strike against), or more than one (*B* for Banned). For workers who are in the *S* and *B* states, it is also important to keep track of the number of days that have elapsed since their last spell of absence. Thus we specify $S_t, t = 1,...,25$ to indicate the *t*'th working day since the end of a spell of absence attracting a single strike, and $B_t, t = 1,...,20$ to indicate the *t*'th working day since the end of a spell of absence attracting a ban. In addition, we need to keep track of whether the worker was absent the day before or not, in order to account correctly for spells. We do this with subscripts *a* and *p*. We adopt the convention that while a worker is absent, he stays in the state he was in before the start of the current spell.

For example, a worker in state F_a has no absence in the previous five weeks other than the spell he is currently in. A worker in state S_p^{12}, has a strike against due to an absence spell that ended 12 days ago, and is not currently experiencing an absence spell. Finally, a worker who is in state B_p^{20}, and who chooses to attend will transit to state S_p^{21}.

This would give 92 states, were it not for the fact that 4 of these states are impossible, and two pairs of states are indistinguishable *ex ante*. The states that are impossible are S_p^1, S_a^2, B_p^1 and B_a^2. This is because the first day of a ban, or one-strike-against, must be preceded by an absence; and the second day must be preceded by an attendance.

The labelling of the states with subscript *a* is a bit misleading, in a couple of cases because during an absence that started with the worker in the clear (for instance), it is not known at the start of each day whether the state should carry the label F_a, or whether it should carry the label S_a^1, since whether the day is a continuation of the current absence spell, or the start of a spell of one-strike-against depends on behaviour on that day. If the worker is absent on a day with this double label, then he continues in F_a; if he attends, he transits to S_p^2. Thus F_a and S_a^1 should not be treated as separate states. For this reason, we use the dual label F_a / S_a^1. Similar remarks apply to any of the other states subscripted *a*, and B_a^1.

This apparently minor twist has important implications for the econometrics, since it implies that it is impossible always to characterise states on the basis of experience (or history) alone. Before one can tell on a particular day whether someone is at the beginning of a ban or not, it is necessary to know whether they are absent on that day. History cannot tell you this, since the ban only starts when a decision to attend is made.

A list of states and possible transitions is given in Figure I below. Each row is labelled with the relevant transition matrix row numbers (in the first column of the Table) and transition matrix column numbers in columns 2 and 3.

Call the generic value function

$$\begin{aligned} E_t^O &= p\left(\sigma_t \le \tilde{\sigma}_t^O\right)\left[E\left\{U\left(0,\sigma_t \mid \sigma_t \le \tilde{\sigma}_t^O\right)\right\} + \delta E_{t+1,p}^{D_1}\right] \\ &\quad + p\left(\sigma_t > \tilde{\sigma}_t^O\right)\left[E\left\{U\left(0,\sigma_t \mid \sigma_t > \tilde{\sigma}_t^O\right)\right\} + \delta E_{t+1,a}^{D_2}\right] \\ &= \int_{-\infty}^{\tilde{\sigma}_t^O} U\left(0,z\right)\phi\left(z\right)dz + \int_{\tilde{\sigma}_t^O}^{\infty} U\left(1,z\right)\phi\left(z\right)dz \\ &\quad + \delta\left[\Phi\left(\tilde{\sigma}_t^O\right)E_{t+1,p}^{D_1} + \left(1-\Phi\left(\tilde{\sigma}_t^O\right)\right)E_{t+1,a}^{D_2}\right] \end{aligned} \tag{5}$$

The first line of (9) represents the contemporaneous expected utility of being in State O ('Origin') for a single period. The second line is the discounted expected value of future optimal policy, where $\Phi(\tilde{\sigma}_t^O)$ and $1-\Phi(\tilde{\sigma}_t^O)$ are the non-zero probabilities of transition into two possible destination states. (In this system, there are only two destination states for each origin state, one (D_1) reached if the worker attends, the other (D_2) reached if the worker is absent. Transitions into all other states have probability zero.)

Expressions for $\boldsymbol{C}$, $\boldsymbol{c}_0$ and $\boldsymbol{\Psi}$ can now be derived. The transition matrix is displayed in Figure II, where $\underline{\Phi}(\tilde{\sigma}_{tp}^s) \equiv 1-\Phi(\tilde{\sigma}_{tp}^s)\forall s$. Since for each origin state O,

$$c_{0O} = \int_{-\infty}^{\tilde{\sigma}_t^O} U(0,z)\phi(z)dz + \int_{\tilde{\sigma}_t^O}^{\infty} U(1,z)\phi(z)dz\ ,$$

$$\frac{\partial c_{0k}}{\partial \tilde{\sigma}^k} = \phi\left(\tilde{\sigma}^k\right)\left[U\left(0,\tilde{\sigma}^k\right) - U\left(1,\tilde{\sigma}^k\right)\right],\ k = 1,\ldots,K \tag{6}$$

Since $\dfrac{\partial c_{jk}}{\partial \tilde{\sigma}^i} = 0$ if $i \neq j$, these derivatives can be arranged into matrix Ψ, the j'th row of which is $\dfrac{\partial c_{jk}}{\partial \tilde{\sigma}^j}; \quad k = 1,...,K$. Assuming the distribution of σ to be rectangular, and with reference to Figure II, this matrix[6] is:

$$
\Psi = \left[\begin{array}{cc|c|cc|ccc|cc|ccc|c|cc|ccccc}
1 & -1 & 0 & 0 & \cdots & \cdots & & \cdots & \cdots & & \cdots & 0 & 0 & 0 & \cdots & \cdots & \cdots & & \cdots & 0 & 0 \\
0 & -1 & 1 & 0 & 0 & \cdots & & & & & & 0 & 0 & \vdots & & & & & & & \vdots \\
\hline
0 & 0 & 0 & 1 & -1 & \cdots & & & & & & 0 & 0 & \vdots & & & & & & & \vdots \\
\hline
\vdots & \vdots & 0 & 0 & 0 & 1 & -1 & 0 & \cdots & & \cdots & 0 & 0 & 0 & 0 & \cdots & \cdots & & \cdots & 0 & 0 \\
\vdots & & 0 & 0 & 0 & 0 & -1 & 0 & & & & & & 1 & 0 & \cdots & \cdots & & \cdots & 0 & 0 \\
\vdots & & \ddots & & 0 & 0 & 0 & \ddots & & & & & \vdots & \vdots & \vdots & \ddots & & & & & \vdots \\
\vdots & & & & \vdots & & & \ddots & \ddots & & & & \vdots & \vdots & \vdots & & & & & & \vdots \\
\vdots & & & \ddots & \vdots & & & & \ddots & & \ddots & & \vdots & \vdots & \vdots & & \ddots & & & & \vdots \\
0 & 0 & & & \vdots & & & & & & \ddots & 1 & -1 & 0 & \vdots & & & & \ddots & & \vdots \\
 & & & & & & & & & & & 0 & -1 & 1 & & & & & & & \\
\hline
1 & -1 & 0 & & & & & & & & & 0 & 0 & 0 & \vdots & & & & & & \vdots \\
0 & -1 & 0 & \cdots & \cdots & \cdots & & \cdots & \cdots & & \cdots & 0 & 0 & 1 & 0 & \cdots & \cdots & & \cdots & 0 & 0 \\
\hline
0 & 0 & \cdots & \cdots & \cdots & \cdots & & \cdots & \cdots & & \cdots & 0 & 0 & 0 & 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\
\hline
 & \vdots & & & & & & & & & & & \vdots & 0 & 0 & 0 & 1 & -1 & & & \vdots \\
 & & & & & & & & & & & & & 1 & 0 & 0 & 0 & -1 & & & \\
 & \vdots & & & & & & & & & & & \vdots & \vdots & \vdots & \vdots & 0 & 0 & \ddots & & \vdots \\
 & & & & & & & & & & & & & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & -1 \\
0 & 0 & \cdots & \cdots & \cdots & \cdots & & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 & -1 \\
\hline
0 & 0 & \cdots & \cdots & \cdots & \cdots & & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & \cdots & \cdots & & \cdots & 0 & 0 \\
0 & 0 & \cdots & \cdots & \cdots & \cdots & & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & \cdots & \cdots & & \cdots & 0 & 0
\end{array}\right]
$$

---

[6] The structure of $\boldsymbol{\Psi}$ is related to that of $\boldsymbol{C}$ in a manner that is worth exploring, partly because it suggests ways in which computation might be made more efficient. The following is true for rectangular σ. Define the real vector, $\boldsymbol{p}$, and the integer vectors, $\boldsymbol{m}$ and $\boldsymbol{n}$:

$$
\begin{aligned}
\boldsymbol{p} &= \left(p_i \mid i \in [1,K]; p_i \in [0,1]\right) \\
\boldsymbol{m} &= \left(m_i \mid i \in [1,K]; m_i \in [1,K]\right) \\
\boldsymbol{n} &= \left(n_i \mid i \in [1,K]; n_i \in [1,K]; m_i \neq n_i \ \forall\, i\right).
\end{aligned}
$$

Now define $\boldsymbol{C}$ as:

$$
\boldsymbol{C} = \left(c_{ij} \mid i,j \in [1,K]; c_{ij} = \left\langle \begin{array}{l} p_i \ \text{if}\, j = m_i \\ 1 - p_i \ \text{if}\, j = n_i \\ 0 \ \text{otherwise} \end{array} \right. \right)
$$

Let $\boldsymbol{P} = \boldsymbol{D}_p$ and $\boldsymbol{Q} = \boldsymbol{I} - \boldsymbol{P}$. Also define matrices $\boldsymbol{M}$ and $\boldsymbol{N}$ as:

$$
\boldsymbol{M} = \left(m_{ij} \mid i,j \in [1,K];\, m_{ij} = \left\langle \begin{array}{l} 1 \ \text{if}\, j = m_i \\ 0 \ \text{otherwise} \end{array} \right. \right)
$$

$$
\boldsymbol{N} = \left(n_{ij} \mid i,j \in [1,K];\, n_{ij} = \left\langle \begin{array}{l} 1 \ \text{if}\, j = n_i \\ 0 \ \text{otherwise} \end{array} \right. \right)
$$

Then $\boldsymbol{C} = \boldsymbol{PM} + \boldsymbol{QN}$ and $\boldsymbol{\Psi} = \boldsymbol{M} - \boldsymbol{N}$.

With the more general Beta distribution, the pattern of zeroes, positives and negatives is retained in this matrix. The non-zero elements are the slopes of the Beta density at each $\tilde{\sigma}_i$. We now have all the expressions required to compute the utility differences using (2) or (6).

Once these are obtained, they are used to solve a vector of equations (7) for optimal morbidity levels in each state. Once these are known, likelihoods of absence and attendance can be calculated. These can be used for two main purposes: evaluation of the incentive properties of the scheme, and construction of likelihood functions, enabling the investigation of worker responses to the scheme. This paper does not undertake to do the second of these in detail, but in the following section I report the results of an evaluation of the Ashby-de-la-Zouche peanut packers attendance incentive scheme.

4. Evaluation of an attendance incentive scheme

The peanut factory attendance incentive scheme seems simple and reasonable. Two spells of absence in last 5 weeks attract an overtime ban of 4 weeks, beginning at the end of the second spell. It turns out, however, that this simple statement hides a multitude of complexities, some of which will be apparent to the reader who has followed me so far. Consider, for instance, the situation of a worker on the day after which a ban starts. For such a worker, the incentive to attend is weakened, because the marginal ban is close to zero. The presence of such perverse incentives in experience-rated schemes is well-known, but I do not think has been consistently analysed before. A second issue that is often ignored in the design of experience-rated schemes is that any scheme will provide incentives not only for or against starting a spell of absence, but also for or against finishing one. The method developed above is able to shed light on both these issues, as we shall shortly see.

I present the evaluation in the form of diagrams, showing for the peanut factory scheme the probabilities of absence for an individual worker with a given Beta distribution of morbidity and discount rate, whose contract specifies given wage, overtime and sick pay rates for given basic hours and overtime hours. Such diagrams can, of course, be generated for any configuration of these parameters. The illustrations below were generated using the following:

Parameter	*Value*
Morbidity distribution: α	0.6
: β	1.6
Discount rate per day (ρ)	0.00007
(Discount factor (δ))	(0.99993)
Wage Rate	200
Overtime Rate	210
Sickpay Rate	100
Total time available	90
Basic Hours	35
Overtime Hours	2

The illustrations are based on two different representations of the attendance control scheme. In one, I have constructed a weekly representation of the rules, which reduces computation time significantly. The other uses the daily representation which leads to the transition matrix, ***C***, described in an earlier section. Later versions of the paper will include more of the latter kind of analysis.

The time-aggregated (weekly) version ignores the issue of current absence, and is probably close to the kind of thinking that is used when these schemes are devised. It has ten states: Clear, One spell of absence *n* weeks ago (*n*=1,...,5); Banned, with *n* weeks to go (*n*=4,...,1). Figure III shows the probabilities of absence generated by the model with the parameter values indicated in the Table above. These probabilities are, of course, conditional on the state occupied. The incentives are quite clearly described. Having one strike against (states 2 – 6) lowers absence immediately by about 1/6, and continues to lower it as the past absence recedes into the past. In state 6, the probability of absence is 2/3 of the probability in state 1. State 7 illustrates a perverse incentive. The absence jumps to about 3%, 20% higher than the 2.4% in state 1. This is because in State 7 the marginal ban drops to zero, so there is effectively no penalty. As the ban proceeds, the marginal ban rises again until the probability of absence in state 10 is much the same as in state 6.

The plot in Figure IV shows the pattern of incentives for the more detailed daily model. The plot is generated in exactly the same way, using the same parameters as before. The 86 states are plotted from left to right in the same order as they are listed in Figure I. State 1 is thus the clear state (present); state 2 is the clear state (absent). It is apparent from the plot that the incentives for attendance are quite different for these two states, since the state 2 absence rate is more than twice the state 1. It is important to note that this is *not* because of any autocorrelation in the stochastic process underlying the model. In this simulation this is iid. The reason for the higher probability of absence among workers in state 2 is that if they attend, they will switch states to one strike against. Indeed, it is clear that there is a similar phenomenon among workers in states 3 – 49. These are states where the worker has one strike against, incurred at various times in the past. Those workers who are in an ongoing spell of absence have weaker incentives to attend than those who are not. This is because on return to work, they will incur a ban. In fact, as their strike against ages, their incentives get increasingly weaker, reaching a high of 7% in state 49. This is because an absence spell started in spell 49 is quite likely not to finish until after the existing strike against expires. By extending their absence spell until the current strike has expired, workers can avoid a ban and continue to work overtime with a new, fresher strike against them. The same remarks apply, with rather less force, to those workers who have a strike against, and who are not in an ongoing absence spell. Starting an absence spell is potentially less expensive for them, because a spell started now can extend into the future to the point where the current strike has expired.

States 50 – 86 are banned states. The incentives to attend are once again stronger for workers who are not in an ongoing spell of absence. Towards the end of a ban, such workers are very keen not to start a new spell, because doing so would trigger a further ban of almost maximum length.

Does time inconsistency matter in the provision of incentives? It appears from the present admittedly limited exercise that it does not much. Figure V, which is in 10 parts, shows the comparison of the 10-state weekly model for a worker who discounts quasihyperbolicly with $\{\delta,\beta\}=\{0.99993,1.0\}$ on the one hand and with $\{\delta,\beta\}=\{0.99993,0.1\}$ on the other. The former discounting is equivalent to the time-consistent exponential regime. The latter discounts the week immediately ahead at 10 times the rate of other weeks. The comparison suggests that time-inconsistency changes optimal decisions, but not by a great deal. The biggest impact is consistently on state 5, where the contrast between the prospects of being one strike against or banned is at its strongest.

6. Conclusion

This paper demonstrates the use of a model which enables the computation of optimal policies for economic agents who are subject to a redemptive experience-rated incentive scheme. I have shown how these optimal policies may be used in evaluating the incentive properties of such schemes. They may also be used to construct likelihoods for data recording the behaviour of such agents. I have not had the opportunity to prepare any illustrations of this use for the model, but this is where the idea that such schemes generate a kind of generalised circular data come to the fore. In this conclusion, I discuss the econometric advantages that circular data provides. These are, as far as I can tell at the moment, three in number.

Firstly, circles have no beginnings. This means that the initial conditions problem can be handled naturally. This is an idea that is pursued at some length in a previous paper (Treble(2007)).

Secondly, as we have seen, circular data enables the investigation of time inconsistency in a constructive way. While this doesn't seem to make a lot of difference to the fortunes of Ashby peanut packers, I can't see any reason why the effects of time inconsistency should generally be small. It may even be that the small differences are the consequence of some special characteristic of the parameter values that I chose for the simulations.

Thirdly, one fairly obvious criticism of the model described above is that the stochastic shocks to health are iid. While this may make some sense in a time-aggregated model, it makes next to no sense in a daily model – everyone knows that if today you started to feel the symptoms of common cold, you will continue to feel them tomorrow, with a very high probability. A topic for future research will be to incorporate time dependence into the stochastic structure of the model. This is worthwhile particularly because it need not introduce an enormous amount of additional complication into the econometrics. After all, circular data were invented to simplify the analysis of autocorrelated systems.

Origin State	**Destination state**	
	If present	**If absent**
1: F_p	**1:** F_p	**2:** F_a / S_a^1
2: F_a / S_a^1	**3:** S_p^2	**2:** F_a / S_a^1
3: S_p^2	**4:** S_p^3	**5:** S_a^3 / B_a^1
4: S_p^3	**6:** S_p^4	**7:** S_a^4 / B_a^1
5: S_a^3 / B_a^1	**50:** B_p^2	**7:** S_a^4 / B_a^1
…	**…**	**…**
46: S_p^{24}	**48:** S_p^{25}	**49:** S_a^{25} / B_a^1
47: S_a^{24} / B_a^1	**50:** B_p^2	**49:** S_a^{25} / B_a^1
48: S_p^{25}	**1:** F_p	**2:** F_a / S_a^1
49: S_a^{25} / B_a^1	**50:** B_p^2	**2:** F_a / S_a^1
50: B_p^2	**51:** B_p^3	**52:** B_a^3 / B_a^1
51: B_p^3	**53:** B_p^4	**54:** B_a^4 / B_a^1
52: B_a^3 / B_a^1	**50:** B_p^2	**54:** B_a^4 / B_a^1
…		
83: B_p^{19}	**85:** B_p^{20}	**86:** B_p^{20} / B_a^1
84: B_a^{27} / B_a^1	**50:** B_p^2	**86:** B_p^{20} / B_a^1
85: B_p^{20}	**40:** S_p^{21}	**41:** S_p^{21} / B_a^1
86: B_p^{20} / B_a^1	**50:** B_p^2	**41:** S_p^{21} / B_a^1

<u>Figure I</u>

	1	2	3	4	5	6	7	⋯	40	41	⋯	48	49	50	51	52
1	$\Phi(\tilde{\sigma}_{tp}^{F})$	$\underline{\Phi}(\tilde{\sigma}_{tp}^{F})$	0	0	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	0	0	⋯	⋯
2	0	$\underline{\Phi}(\tilde{\sigma}_{ta}^{F})$	$\Phi(\tilde{\sigma}_{ta}^{F})$	0	0	⋯	⋯	⋯	⋯							
3	0	0	0	$\Phi(\tilde{\sigma}_{tp}^{S_2})$	$\underline{\Phi}(\tilde{\sigma}_{tp}^{S_2})$	⋯	⋯	⋯								
4	⋮	⋮	0	0	0	$\Phi(\tilde{\sigma}_{tp}^{S_3})$	$\underline{\Phi}(\tilde{\sigma}_{tp}^{S_3})$	0	⋯	⋯	⋯	⋯	0	0	0	⋯
5	⋮		0	0	0	0	$\underline{\Phi}(\tilde{\sigma}_{ta}^{S_3})$	0					⋮	$\Phi(\tilde{\sigma}_{ta}^{S_3})$	0	
	⋮		⋱		0	0	0	⋱					⋮	⋮	⋮	⋱
	⋮			⋱	⋮			⋱					⋮	⋮	⋮	
	⋮				⋮						⋱		⋮	⋮	⋮	
46	⋮				⋮						⋱	$\Phi(\tilde{\sigma}_{tp}^{S_{24}})$	$\underline{\Phi}(\tilde{\sigma}_{tp}^{S_{24}})$	0	⋮	
47	0	0			⋮							0	$\underline{\Phi}(\tilde{\sigma}_{ta}^{S_{24}})$	$\Phi(\tilde{\sigma}_{ta}^{S_{24}})$	⋮	
48	$\Phi(\tilde{\sigma}_{tp}^{S_{25}})$	$\underline{\Phi}(\tilde{\sigma}_{tp}^{S_{25}})$	0									0	0	0	⋮	
49	0	$\underline{\Phi}(\tilde{\sigma}_{ta}^{S_{25}})$	0	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	0	0	$\Phi(\tilde{\sigma}_{ta}^{S_{25}})$	0	
50	0	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	0	0	$\Phi(\tilde{\sigma}_{tp}^{B_2})$	$\underline{\Phi}(\tilde{\sigma}_{tp}^{B_2})$
51	⋮												⋮	0	0	0
52	⋮												⋮	$\Phi(\tilde{\sigma}_{ta}^{B_3})$	0	0
	⋮												⋮	⋮		
83	⋮										0	⋯	0	0	0	0
84	0	⋯	⋯	⋯	⋯	⋯	⋯	0	0	0	0	⋯	0	$\Phi(\tilde{\sigma}_{ta}^{B_{19}})$	0	⋯
85	0	⋯	⋯	⋯	⋯	⋯	⋯	0	$\Phi(\tilde{\sigma}_{tp}^{B_{20}})$	$\underline{\Phi}(\tilde{\sigma}_{tp}^{B_{20}})$	0	⋯	0	⋮	0	⋯
86	0	⋯	⋯	⋯	⋯	⋯	⋯	0	0	$\underline{\Phi}(\tilde{\sigma}_{ta}^{B_{20}})$	0	⋯	0	$\Phi(\tilde{\sigma}_{ta}^{B_{20}})$	0	⋯

Figure II: The transition matrix

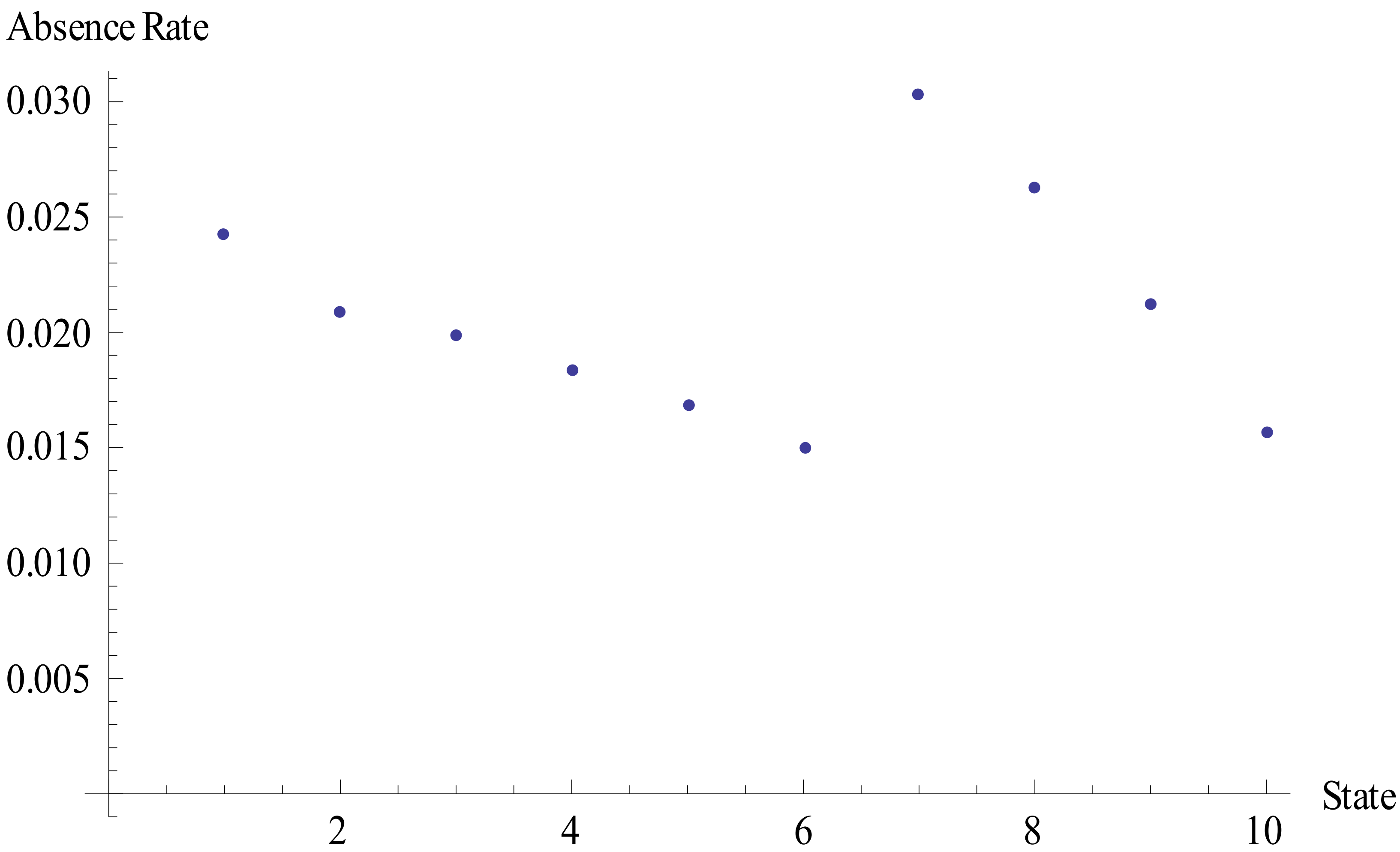

Figure III

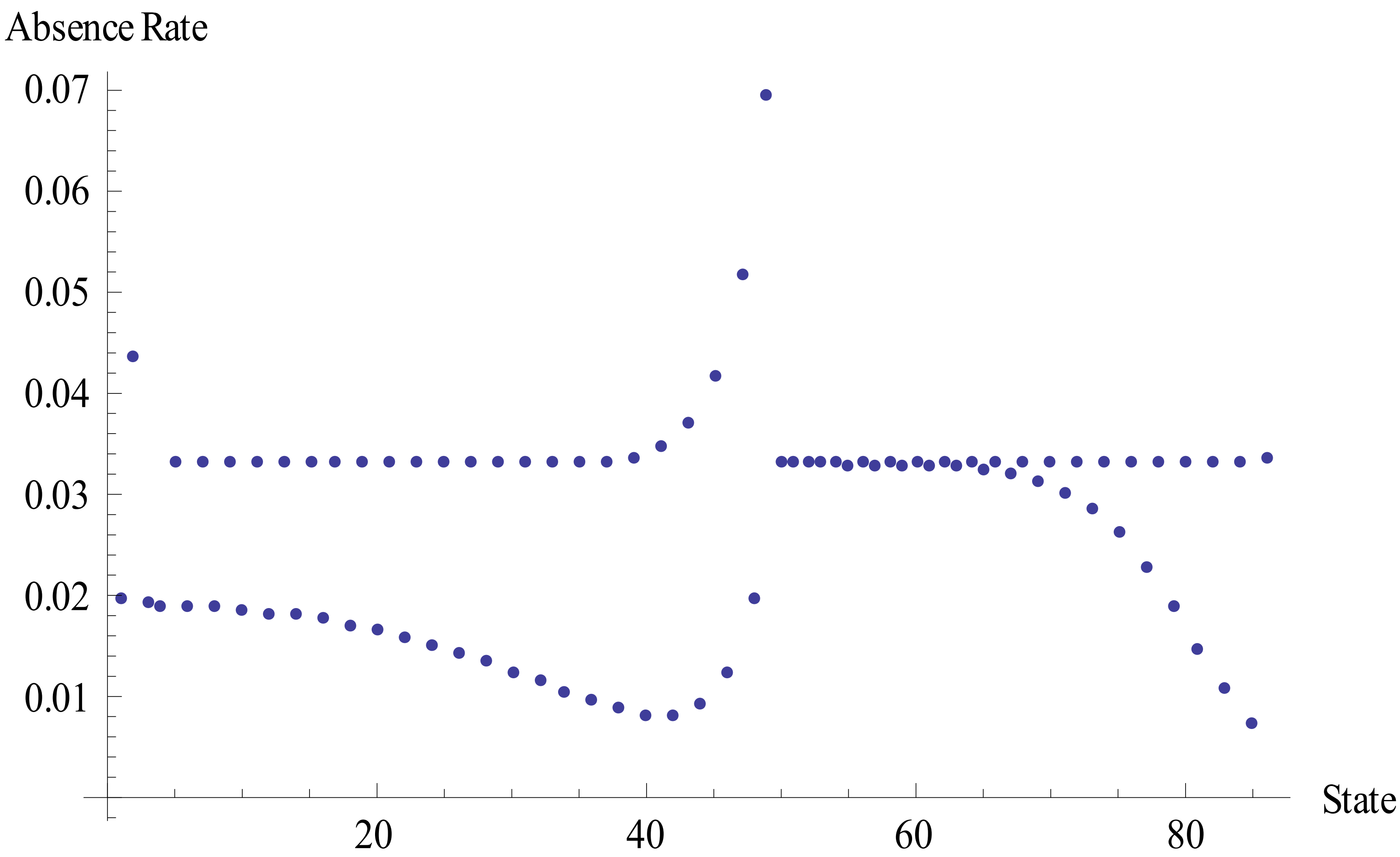

Figure IV

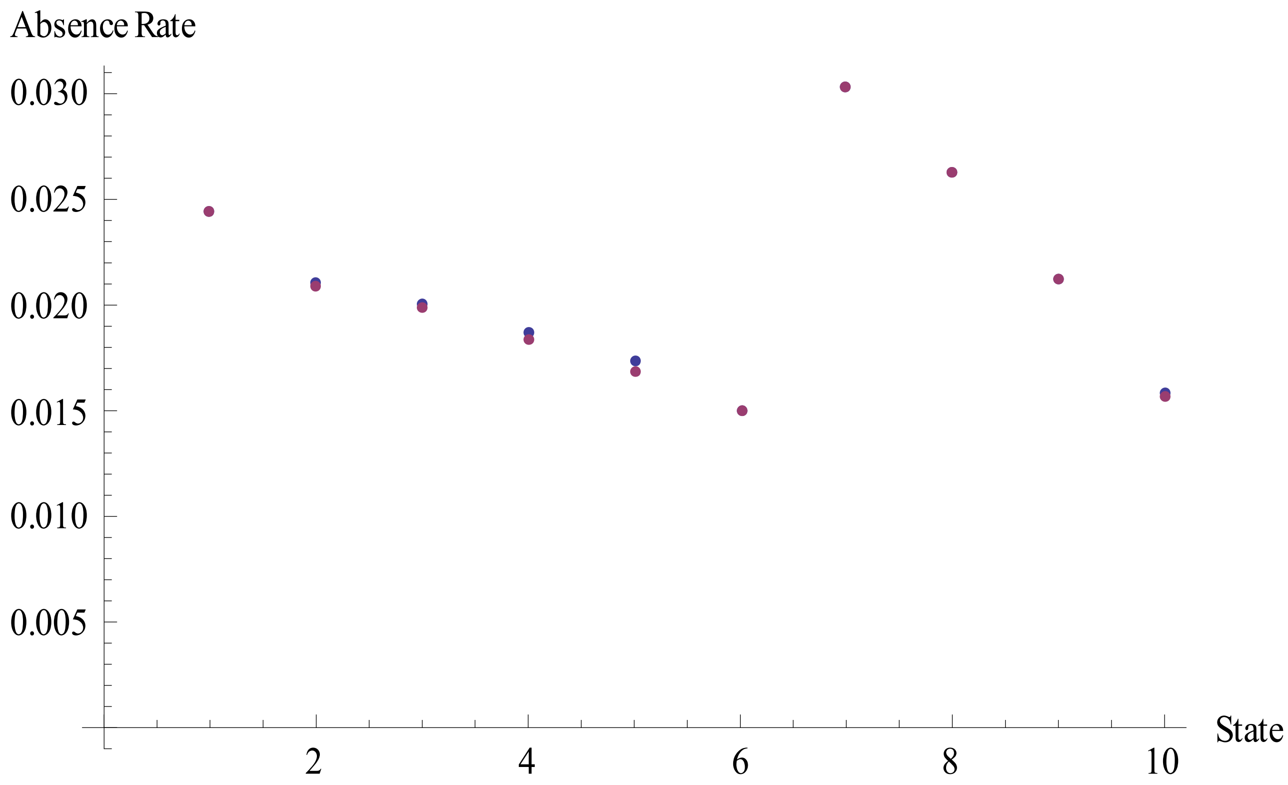

Figure V:State 1: Optimal policies under exponential (blue) and quasihyperbolic (kind of pinky) discounting

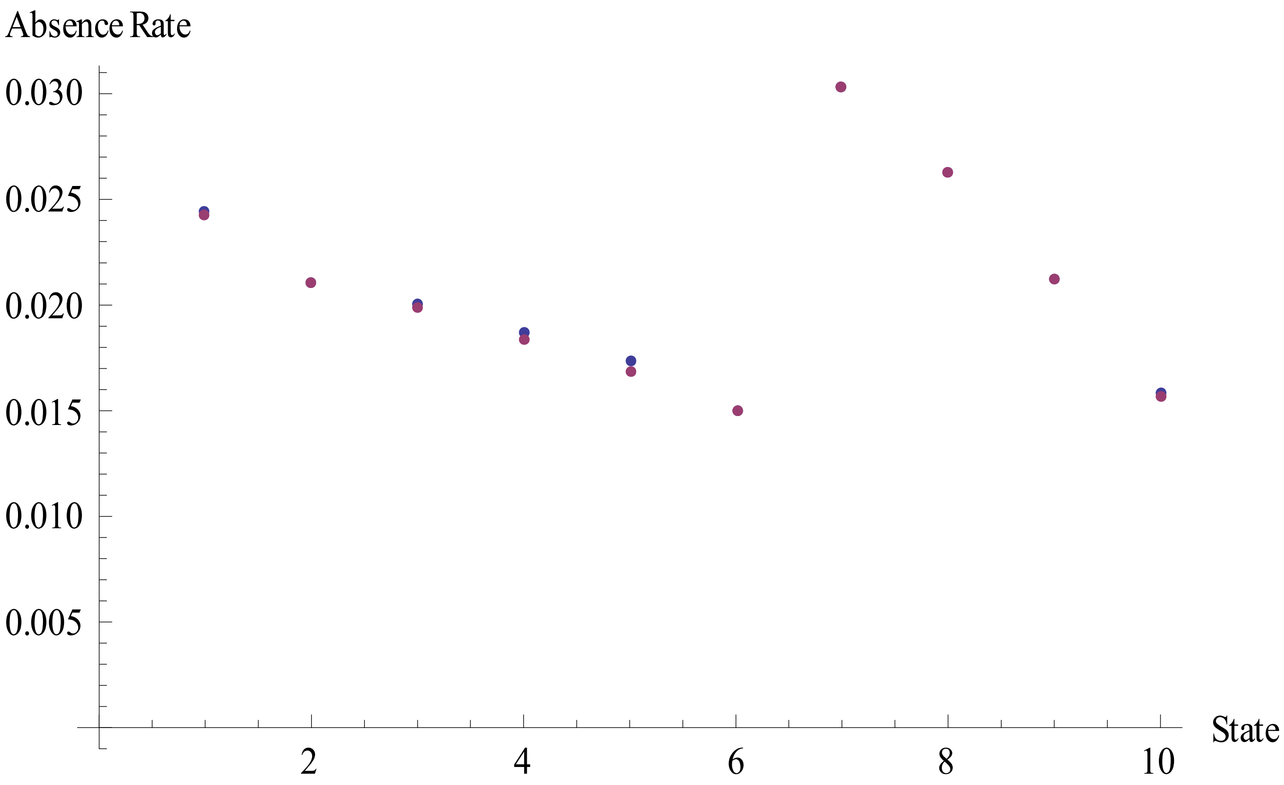

Figure V: State 2: Optimal policies under exponential (blue) and quasihyperbolic (kind of pinky) discounting

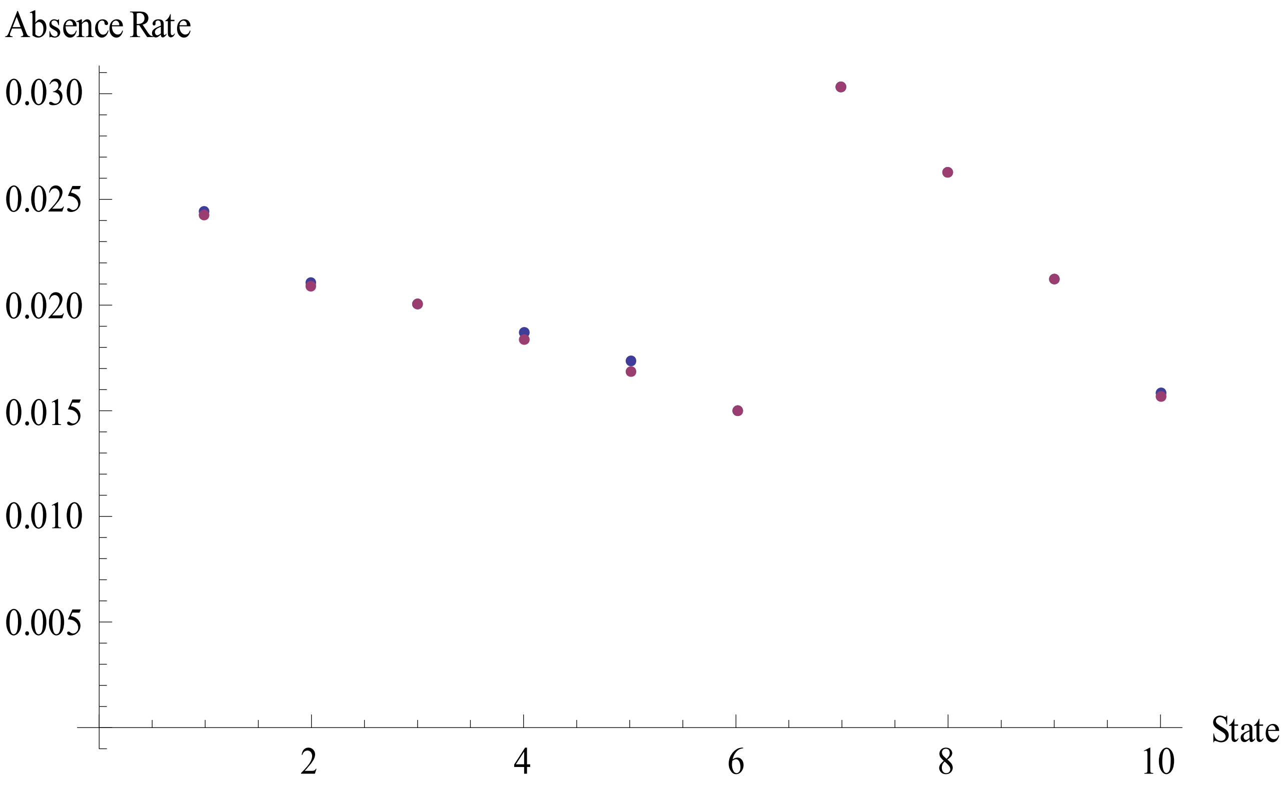

Figure V: State 3: Optimal policies under exponential (blue) and quasihyperbolic (kind of pinky) discounting

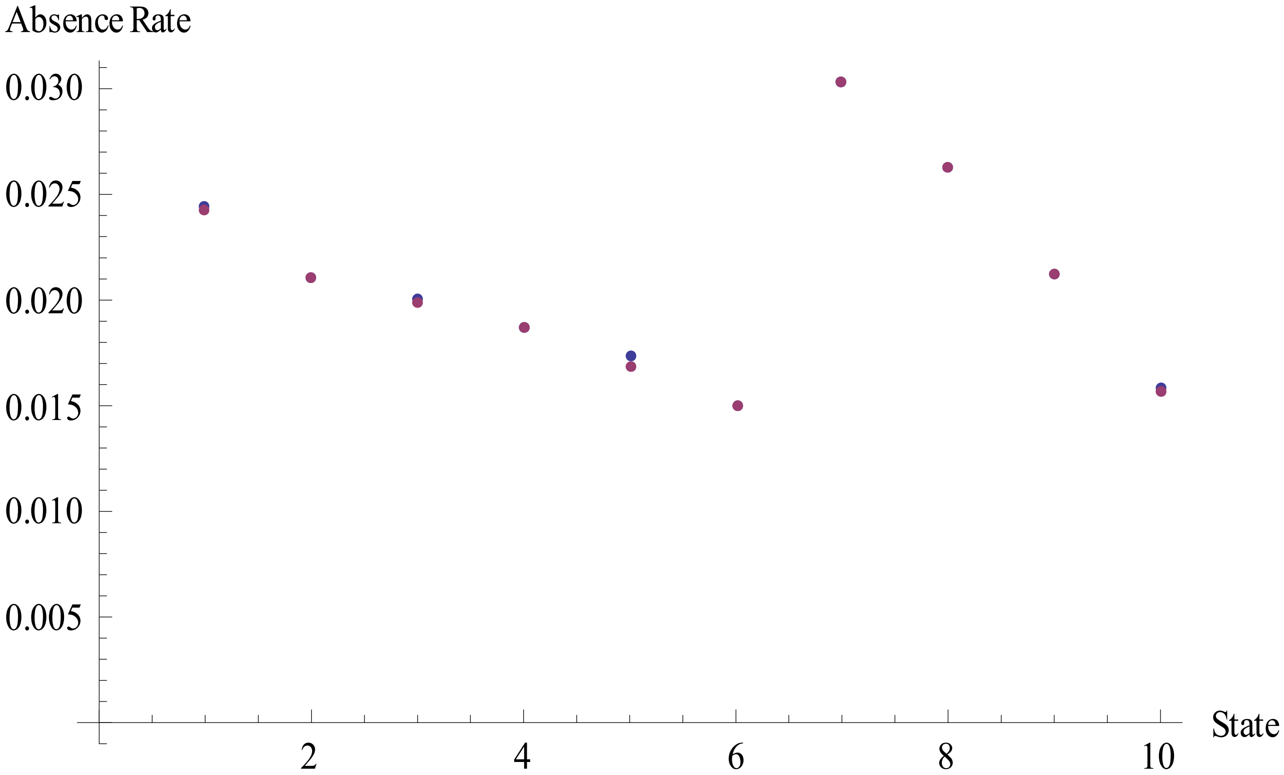

Figure V: State 4: Optimal policies under exponential (blue) and quasihyperbolic (kind of pinky) discounting

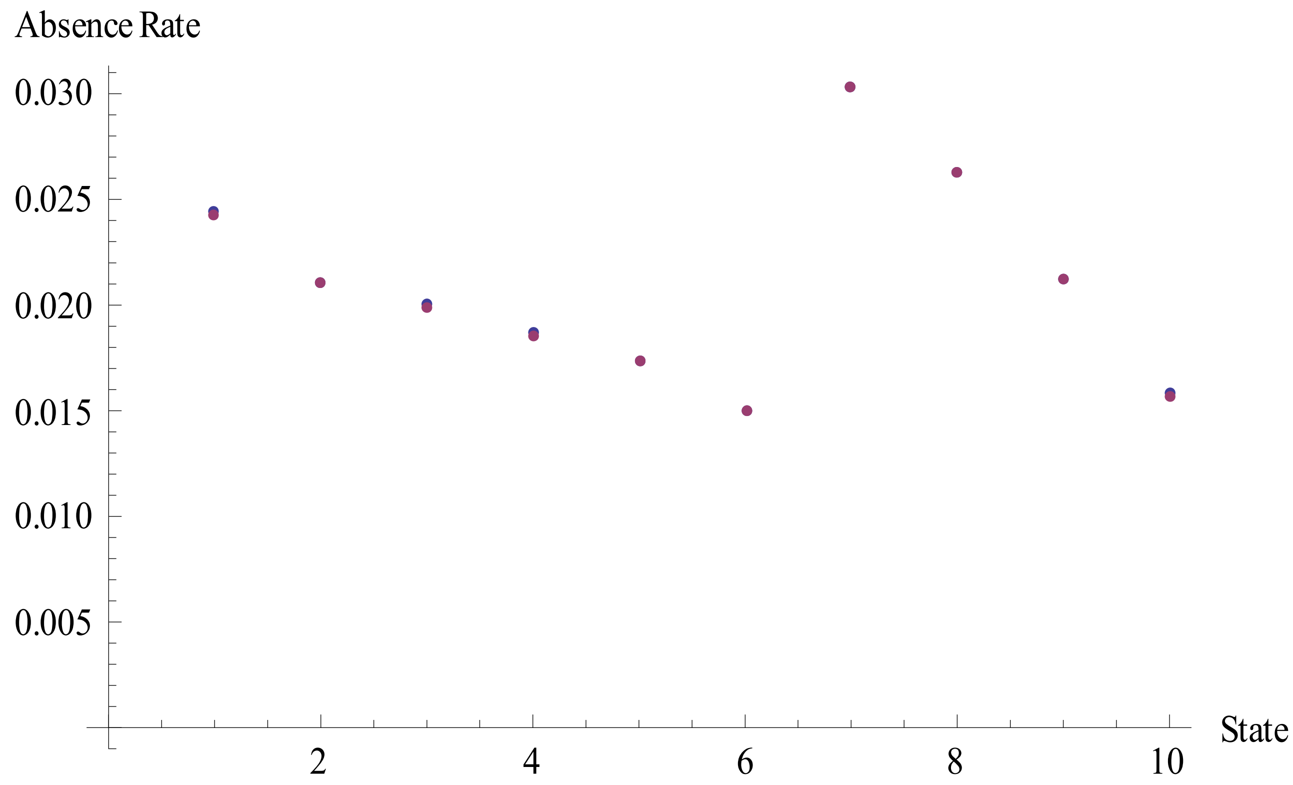

Figure V: State 5: Optimal policies under exponential (blue) and quasihyperbolic (kind of pinky) discounting

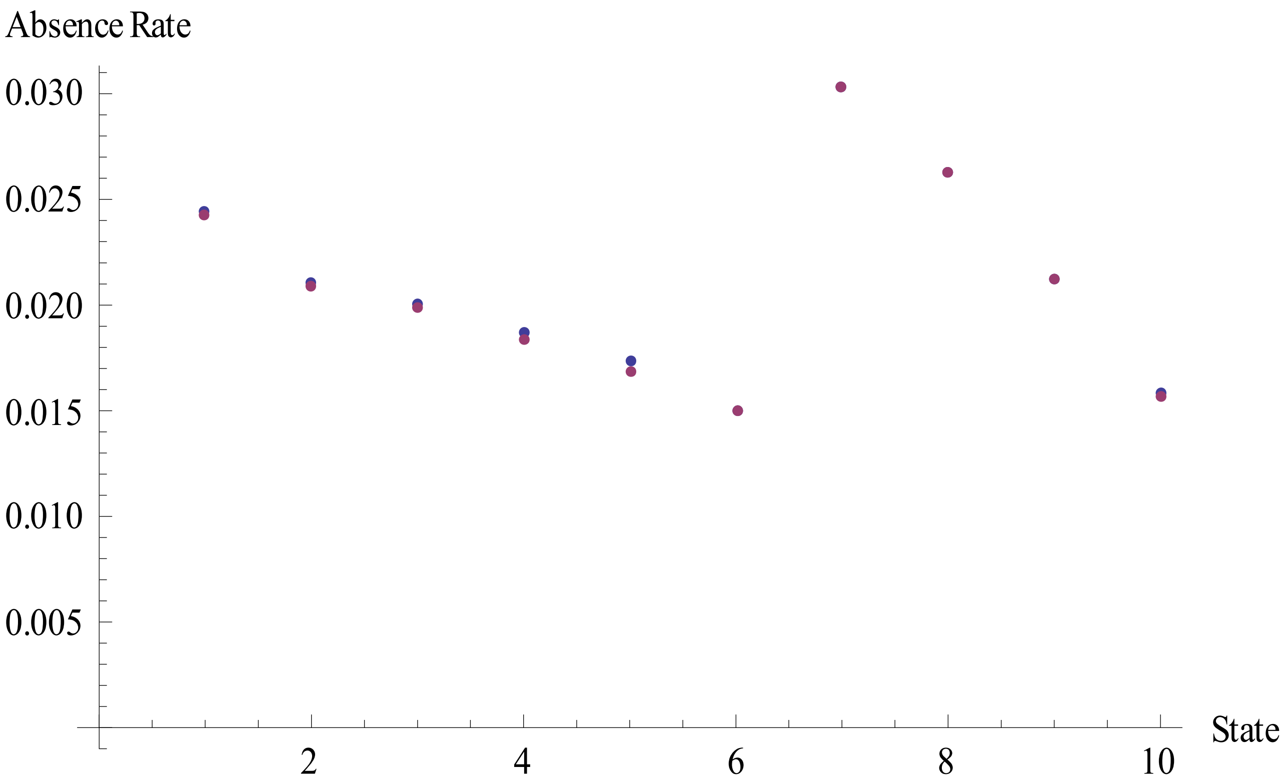

Figure V: State 6: Optimal policies under exponential (blue) and quasihyperbolic (kind of pinky) discounting

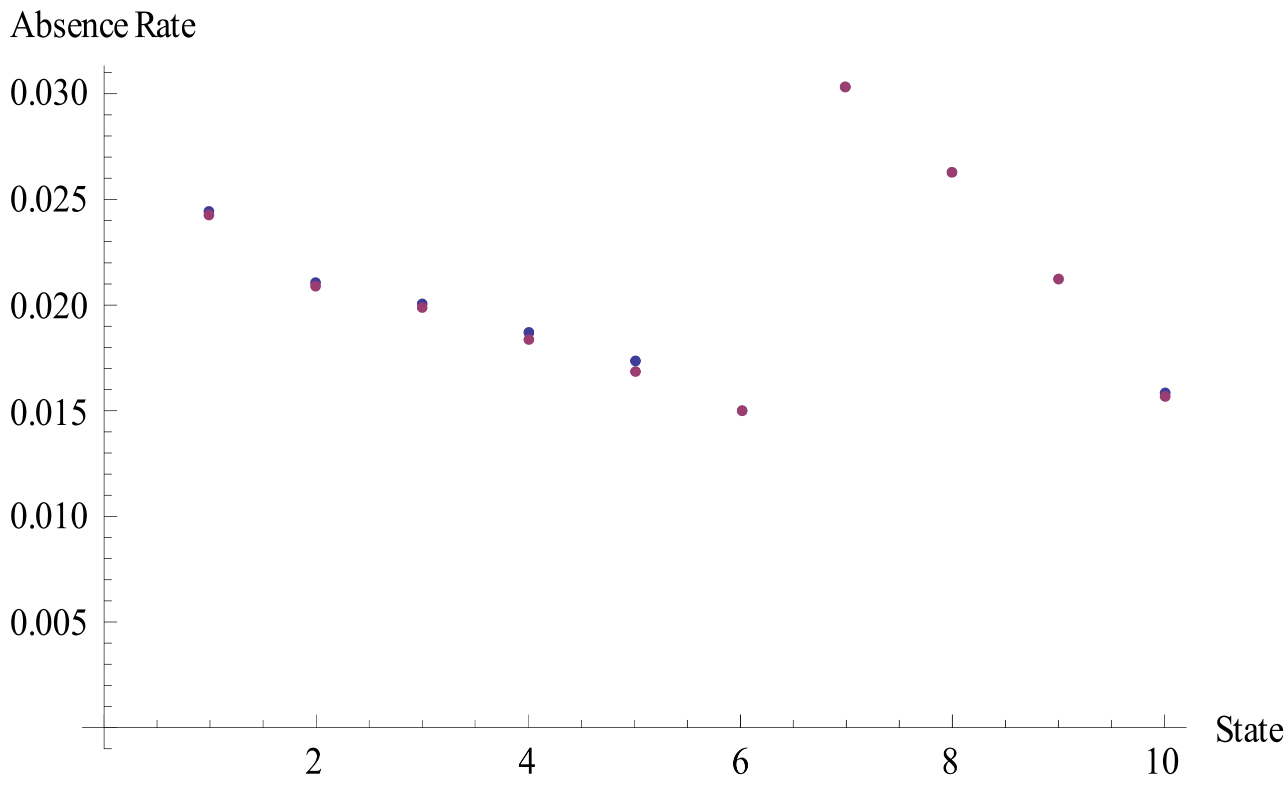

Figure V: State 7: Optimal policies under exponential (blue) and quasihyperbolic (kind of pinky) discounting

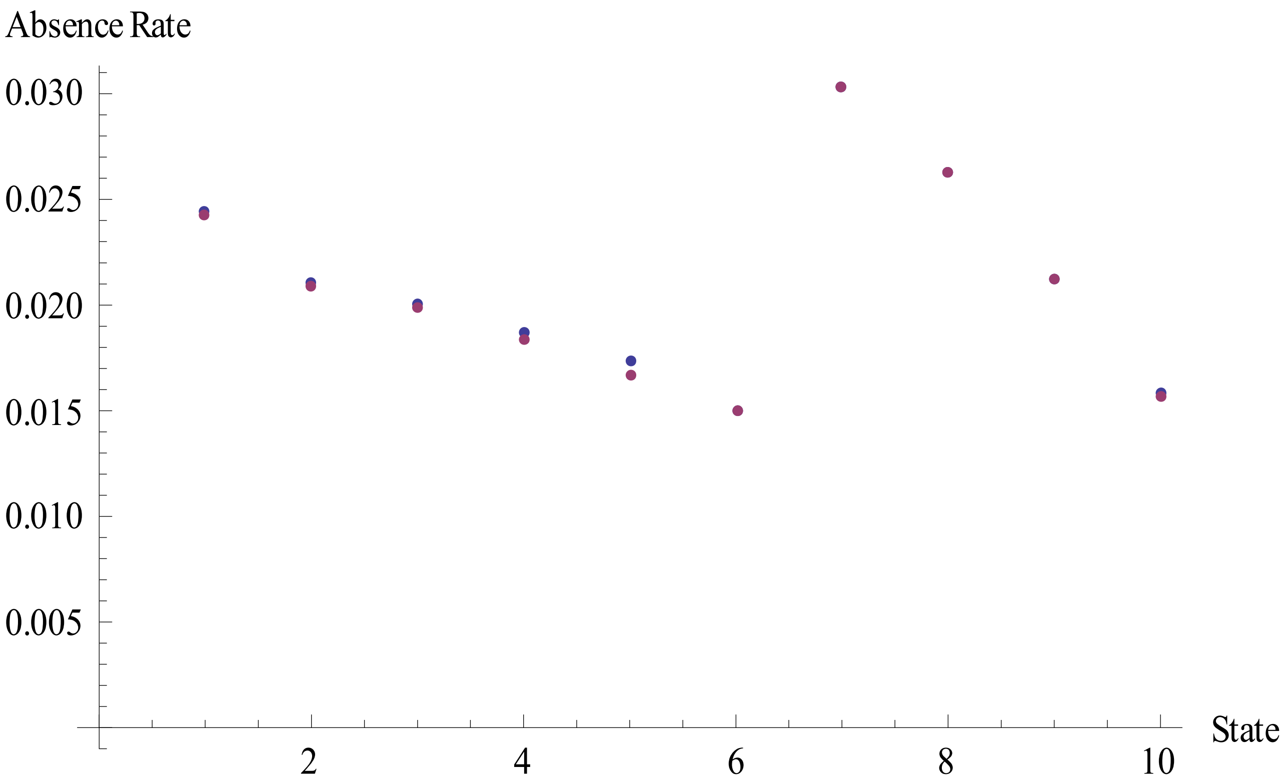

Figure V: State 8: Optimal policies under exponential (blue) and quasihyperbolic (kind of pinky) discounting

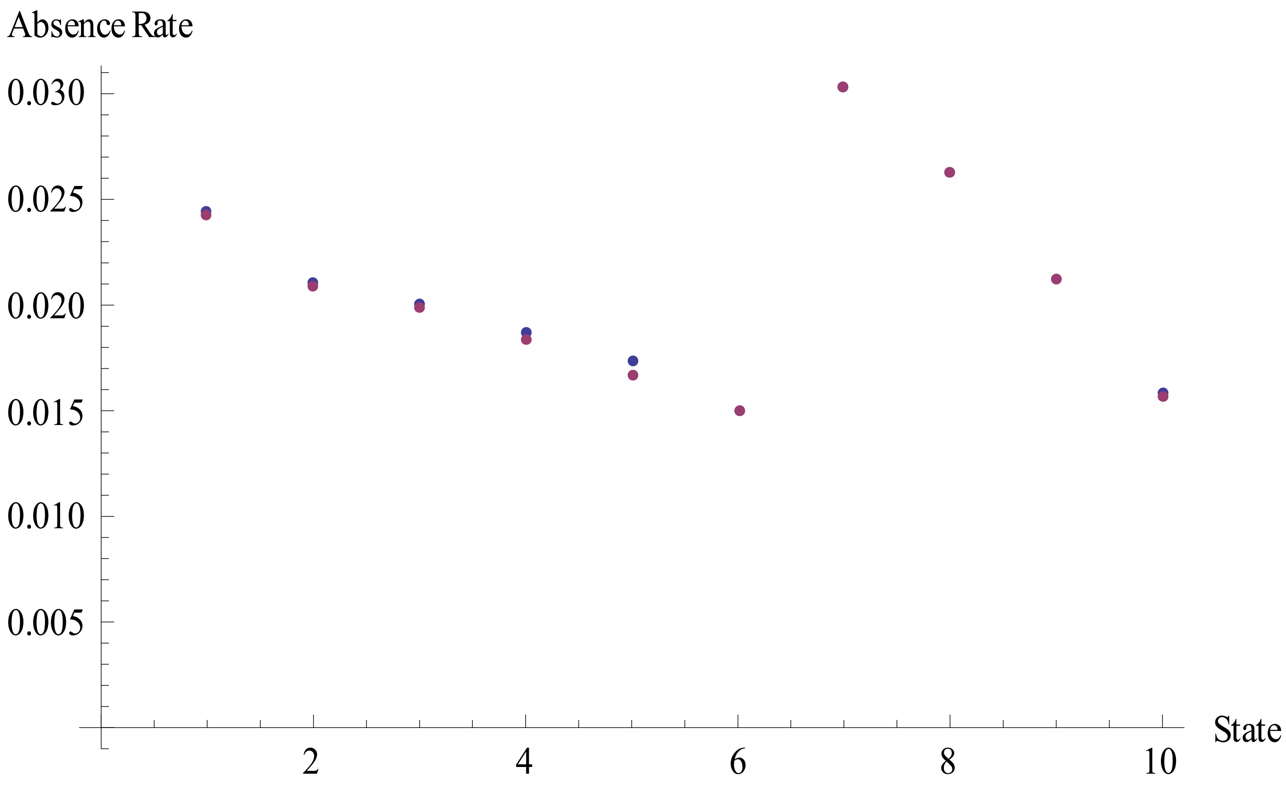

Figure V: State 9: Optimal policies under exponential (blue) and quasihyperbolic (kind of pinky) discounting

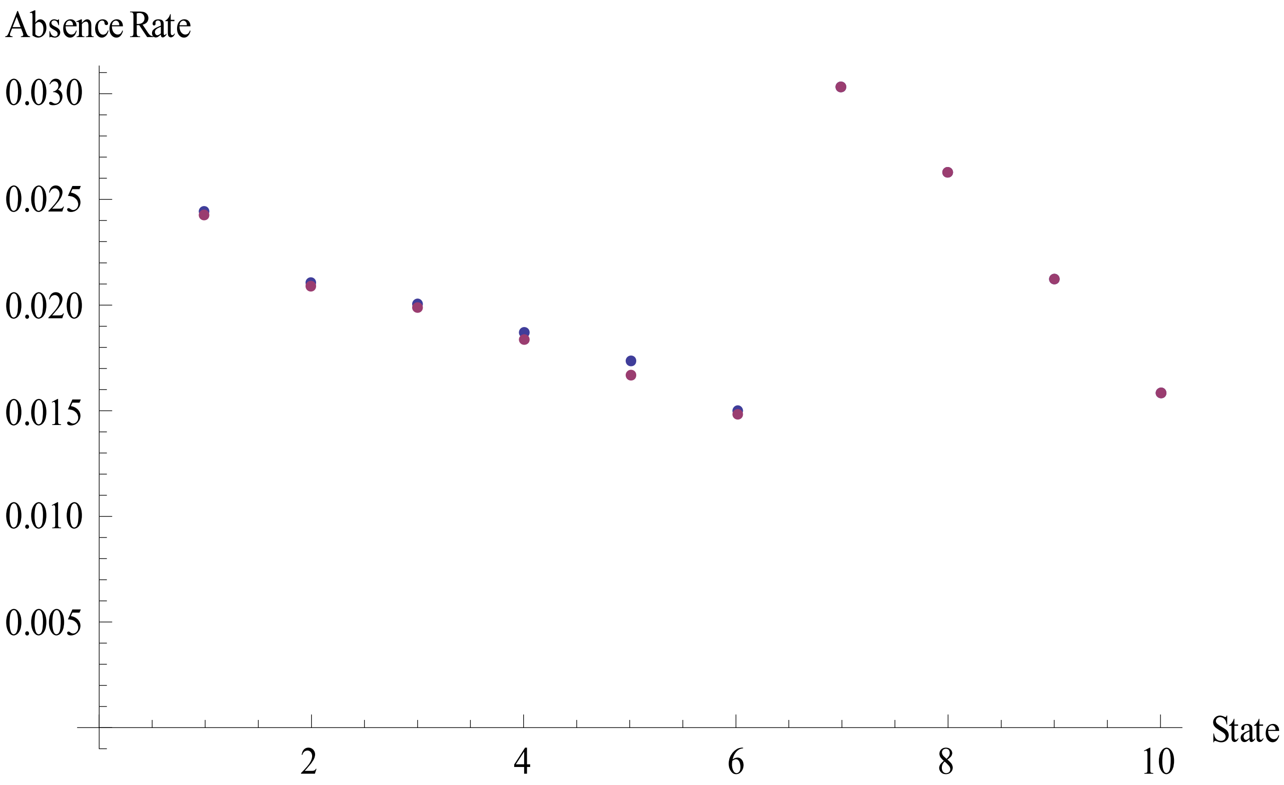

Figure V: State 10: Optimal policies under exponential (blue) and quasihyperbolic (kind of pinky) discounting

References

R. L. Anderson (1942): “Distribution of the Serial Correlation Coefficient” *The Annals of Mathematical Statistics*, Vol. 13, No. 1. pp. 1-13.

T. A. Barmby, C. D. Orme, J. G. Treble (1991): “Worker Absenteeism: An Analysis Using Microdata” *The Economic Journal*, Vol. 101, No. 405. pp. 214-229.

D. R. Cox (1958): “The Regression Analysis of Binary Sequences” *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 20, No. 2. , pp. 215-242.

J. Durbin (1984): “Present Position and Potential Developments: Some Personal Views: Time Series Analysis” *Journal of the Royal Statistical Society. Series A (General)*, Vol. 147, No. 2, The 150th Anniversary of the Royal Statistical Society., pp. 161-173.

D. Fudenberg and J. Tirole (2000): “Customer Poaching and Brand Switching”, *The RAND Journal of Economics*, vol. 31, n. 4, p. 634-657.

W.H.J. Hassink and P. Köning (2005): “Do financial bonuses to employees reduce their absenteeism? Outcome of a lottery.” IZA Discussion Paper No. 1644, Bonn.

P. Klemperer (1995): “Competition when Consumers have Switching Costs: An Overview with Applications to Industrial Organization, Macroeconomics, and International Trade,” *Review of Economic Studies*, 62, 515-539.

D. Laibson (1997) “Golden Eggs and Hyperbolic Discounting,” *Quarterly Journal of Economics*, 62, May, pp. 443-77.

J.G.Treble (2007): “Circularity and the Initial Conditions Problem in Dynamic Binary Choice Models” available at http://www.swan.ac.uk/economics/staff/jgt2.htm

J.G. Treble, T.A. Barmby, S. Brown (2006): “Discrete Choice in a Dynamic Model with Endogenous Constraints; or How to Make a Cheaper Peanut”, mimeo (2006).

G. S. Watson; J. Durbin (1951): “Exact Tests of Serial Correlation using Noncircular Statistics” *The Annals of Mathematical Statistics*, Vol. 22, No. 3., pp. 446-451.

Appendix I

The following results are used later:

Proposition A1: Let $\boldsymbol{X}$ be a matrix with elements $x_{ij}, i, j = 1,...,K$, and let $\kappa_X(i,j)$ denote the cofactor of x_{ij} in $\boldsymbol{X}$. Then

a) $\dfrac{\partial |\boldsymbol{X}|}{\partial x_{ij}} = \kappa_X(i,j)$

b) $\dfrac{\partial |\boldsymbol{X}|}{\partial \boldsymbol{X}} = adj'(\boldsymbol{X}) = |\boldsymbol{X}|\left(\boldsymbol{X}^{-1}\right)'$.

Proof: a) Expanding along the *i*'th row of $\boldsymbol{X}$, gives $|\boldsymbol{X}| = \sum_{k=1}^{K} x_{ik}\kappa_X(i,k)$. Since x_{ij} doesn't appear in the cofactor of any element in the *i*'th row, the result follows.

b) $\boldsymbol{X}^{-1} = \dfrac{adj(\boldsymbol{X})}{|\boldsymbol{X}|}$ and $adj(\boldsymbol{X})$ is the transpose of the matrix of cofactors of elements of $\boldsymbol{X}$. From part a) the result follows. □

Proposition A2: Let $\boldsymbol{X}_{\bullet n=y}$ denote the matrix $\boldsymbol{X}$ with its *n*'th column replaced by the vector $\boldsymbol{y}$, and $\boldsymbol{X}_{n\bullet=y}$ denote the matrix $\boldsymbol{X}$ with its *n*'th row replaced by the vector $\boldsymbol{y}$.

a) $\dfrac{\partial |\boldsymbol{X}_{\bullet n=y}|}{\partial x_{ij}} = \begin{cases} \kappa_{X_{\bullet n=y}}(i,j) \text{ if } j \neq n \\ 0 \text{ otherwise} \end{cases}$

b) $\dfrac{\partial |\boldsymbol{X}_{\bullet n=y}|}{\partial \boldsymbol{X}} = \left(adj'\left(\boldsymbol{X}_{\bullet n=y}\right)\right)_{\bullet n=y} = |\boldsymbol{X}_{\bullet n=y}|\left(\left(\boldsymbol{X}_{\bullet n=y}^{-1}\right)'\right)_{\bullet n=y} = |\boldsymbol{X}_{\bullet n=y}|\left(\left(\boldsymbol{X}_{\bullet n=y}^{-1}\right)_{n\bullet=y}\right)'$

c) $\dfrac{\partial |\boldsymbol{X}_{\bullet n=y}|}{\partial y_i} = \kappa_X(i,n)$

Proof: Expanding along the *j*'th column of $\boldsymbol{X}_{\bullet n=y}$, gives

$$\left|\boldsymbol{X}_{\bullet n=y}\right| = \begin{cases} \sum_{k=1}^{K} x_{kj}\kappa_{X_{\bullet n=y}}\left(k, j\right) \text{ for } j \neq n \\ \sum_{k=1}^{K} y_k\kappa_{X_{\bullet n=y}}\left(k, j\right) \text{ for } j = n \end{cases}.$$

a) Since x_{kj} doesn't appear in the cofactor of any element in the *j*'th column,

$$\frac{\partial\left|\boldsymbol{X}_{\bullet n=y}\right|}{\partial x_{kj}} = \begin{cases} \kappa_{X_{\bullet n=y}}\left(k, j\right) \text{ for } j \neq n \\ 0 \text{ for } j = n \end{cases}$$

b) The first equality follows from similar reasoning as given for Proposition A1, part a). The remaining equalities follow from the definitions of $\boldsymbol{X}^{-1}$, $\boldsymbol{X}_{\bullet n=y}$ and $\boldsymbol{X}_{n\bullet=y}$.

c) Follows directly from the definition of the determinant above. □

Definition A1: An *elementary matrix* $\boldsymbol{E}_{ij}$ is a $J \times K$ matrix of zeroes, except for the *j*,*k*'th element which is equal to 1.

Definition A2: An *elementary vector* $\boldsymbol{e}_j$ is an *J*-dimensional column vector of zeroes, except for the *j*'th element which is equal to 1. Note that $\boldsymbol{e}_j^{'}\boldsymbol{e}_k = 1$ and $\boldsymbol{e}_j\boldsymbol{e}_k^{'} = \boldsymbol{E}_{jk}$. Elementary vectors are also useful for extracting rows or columns of matrices, or for constructing them from elements of the matrix. For instance, $\boldsymbol{X}_{i\bullet}$, the *i*'th row of matrix ***X***, can be written as column vector: $\boldsymbol{X}_{j\bullet} = \boldsymbol{X}^{'}\boldsymbol{e}_j = \sum_k x_{jk}\boldsymbol{e}_k$.

Proposition A3:

The trace of a square matrix can be written as $tr\left(\boldsymbol{X}\right) = \sum_j \boldsymbol{X}_{j\bullet}^{'}\boldsymbol{e}_j$.

Proof: $\boldsymbol{X}_{j\bullet} = \left\{x_{j1}, x_{j2}, ..., x_{jj}, ..., x_{jK}\right\}^{'}$. Postmultiplying by $\boldsymbol{e}_j$, gives scalar x_{jj}. The trace is the sum of these diagonal elements.

Proposition A4:

Let square matrix $\boldsymbol{Y}$ be a function of square matrix $\boldsymbol{X}$, both of order K. Then

a) [Jacobi's Formula]: $\frac{\partial|\boldsymbol{Y}|}{\partial x_{ij}} = tr\left[\left(\frac{\partial \boldsymbol{Y}}{\partial x_{ij}}\right)' \mathrm{K}^{\boldsymbol{Y}}\right]$, where $\mathrm{K}^{\boldsymbol{Y}}$ denotes the matrix of cofactors of $\boldsymbol{Y}$; and

b) $\frac{\partial|\boldsymbol{Y}|}{\partial \boldsymbol{X}} = \sum_{i=1}^{K}\sum_{j=1}^{K} \boldsymbol{E}_{ij} \frac{\partial|\boldsymbol{Y}|}{\partial x_{ij}}$.

Proof: b) The second part of the proposition follows immediately from the definitions of the derivative of a scalar with respect to a matrix, and of $\boldsymbol{E}_{ij}$, so it is necessary only to prove the first part.

a) By the chain rule and Proposition A1.1,

$$\begin{aligned}
\frac{\partial|\boldsymbol{Y}|}{\partial x_{ij}} &= \sum_{r=1}^{K}\sum_{s=1}^{K} \frac{\partial|\boldsymbol{Y}|}{\partial y_{rs}} \frac{\partial y_{rs}}{\partial x_{ij}} = \sum_{r=1}^{K}\sum_{s=1}^{K} \boldsymbol{\kappa}_{\boldsymbol{Y}}(r,s) \frac{\partial y_{rs}}{\partial x_{ij}} \\
&= \sum_{r=1}^{K}\sum_{s=1}^{K} \boldsymbol{\kappa}_{\boldsymbol{Y}}(r,s) \frac{\partial y_{rs}}{\partial x_{ij}} \boldsymbol{e}_r' \boldsymbol{e}_s = \sum_{r=1}^{K}\sum_{s=1}^{K} \boldsymbol{\kappa}_{\boldsymbol{Y}}(r,s) \boldsymbol{e}_r' \frac{\partial y_{rs}}{\partial x_{ij}} \boldsymbol{e}_s \\
&= \sum_{r=1}^{K}\left(\sum_{s=1}^{K} \boldsymbol{\kappa}_{\boldsymbol{Y}}(r,s) \boldsymbol{e}_r' \frac{\partial y_{rs}}{\partial x_{ij}} \boldsymbol{e}_s\right) = \sum_{r=1}^{K}\left(\left(\mathrm{K}_{r\bullet}^{\boldsymbol{Y}}\right)'\left(\frac{\partial Y}{\partial x_{ij}}\right)_{r\bullet}\right) \\
&= tr\left[\mathrm{K}^{\boldsymbol{Y}}\left(\frac{\partial Y}{\partial x_{ij}}\right)'\right] = tr\left[\left(\frac{\partial Y}{\partial x_{ij}}\right)' \mathrm{K}^{\boldsymbol{Y}}\right]
\end{aligned}$$

The first equality in the penultimate line follows because $\boldsymbol{\kappa}_{\boldsymbol{Y}}(r,s) \boldsymbol{e}_r' \frac{\partial y_{rs}}{\partial x_{ij}} \boldsymbol{e}_s = 0$ if $r \neq s$.

The last line follows because the r'th diagonal element of the product $\boldsymbol{UV}'$ is the r'th row of $\boldsymbol{U}$ postmultiplied by the r'th column of $\boldsymbol{V}'$ (the transpose of the r'th row of $\boldsymbol{V}$). □

Proof of Proposition I:

The vector $\boldsymbol{EV} = \{EV_n \mid n = 1,...,K\}$ can be written as a function of the c's alone. Equation **(8)** implies:

$$\begin{pmatrix} EV_1 \\ EV_2 \\ . \\ . \\ . \\ EV_K \end{pmatrix} = \begin{pmatrix} c_{01} \\ c_{02} \\ . \\ . \\ . \\ c_{0K} \end{pmatrix} + \delta \begin{bmatrix} c_{11} & c_{12} & . & . & . & c_{1K} \\ c_{21} & & & & & . \\ . & & & & & . \\ . & & & & & . \\ . & & & & & . \\ c_{K1} & . & . & . & . & c_{KK} \end{bmatrix} \begin{pmatrix} EV_1 \\ EV_2 \\ . \\ . \\ . \\ EV_K \end{pmatrix}$$

or

$$\boldsymbol{EV} = \boldsymbol{c}_0 + \delta \boldsymbol{C}.\boldsymbol{EV} \qquad \text{A(7)}$$

Thus $\boldsymbol{EV} = (\boldsymbol{I} - \delta\boldsymbol{C})^{-1} \boldsymbol{c}_0$ and EV_n can be computed using Cramer's Rule. Let $\boldsymbol{A} = \boldsymbol{I} - \delta\boldsymbol{C}$. Then $EV_n = \dfrac{|A_{\bullet n = c_0}|}{|A|}$.

Now, for each *i,*

$$\frac{\partial EV_n}{\partial \tilde{\sigma}_i} = \sum_{j=0}^{K}\sum_{k=1}^{K} \frac{\partial EV_n}{\partial c_{jk}} \frac{dc_{jk}}{d\tilde{\sigma}_i} = 0$$

$$\Rightarrow \sum_{k=1}^{K} \frac{\partial EV_n}{\partial c_{0k}} \frac{dc_{0k}}{d\tilde{\sigma}_i} = -\sum_{j=1}^{K}\sum_{k=1}^{K} \frac{\partial EV_n}{\partial c_{jk}} \frac{dc_{jk}}{d\tilde{\sigma}_i}$$

$$\Rightarrow \frac{\partial EV_n}{\partial c_{0i}} \frac{dc_{0i}}{d\tilde{\sigma}_i} = -\sum_{j=1}^{K}\sum_{k=1}^{K} \frac{\partial EV_n}{\partial c_{jk}} \frac{dc_{jk}}{d\tilde{\sigma}_i} \text{ since } \frac{dc_{0k}}{d\tilde{\sigma}_i} = 0 \text{ for } k \neq i$$

$$\Rightarrow \frac{\partial EV_n}{\partial c_{0i}} \frac{dc_{0i}}{d\tilde{\sigma}_i} = -\sum_{k=1}^{K} \frac{\partial EV_n}{\partial c_{ik}} \frac{dc_{ik}}{d\tilde{\sigma}_i} \text{ since } \frac{dc_{jk}}{d\tilde{\sigma}_i} = 0 \text{ for } j \neq i$$

Write this in matrix form as

$$diag\left[\lambda \psi'\right] = -diag\left[\Lambda \Psi'\right] \qquad \text{A(8)}$$

where λ and ψ are the K-dimensional column vectors:

$$\lambda = \left\{ \frac{\partial EV_n}{\partial c_{0i}}, i = 1,...,K \right\} \text{ and } \psi = \left\{ \frac{dc_{0i}}{d\tilde{\sigma}_i}, i = 1,...,K \right\}$$

and Λ and Ψ are the $K \times K$ matrices:

$$\Lambda = \begin{bmatrix} \frac{\partial EV_n}{\partial c_{11}} & \frac{\partial EV_n}{\partial c_{12}} & \cdots & \frac{\partial EV_n}{\partial c_{1K}} \\ \frac{\partial EV_n}{\partial c_{21}} & & & \vdots \\ \vdots & & & \vdots \\ \frac{\partial EV_n}{\partial c_{K1}} & \cdots & \cdots & \frac{\partial EV_n}{\partial c_{KK}} \end{bmatrix} \text{ and } \Psi = \begin{bmatrix} \frac{\partial c_{11}}{\partial \tilde{\sigma}_1} & \frac{\partial c_{12}}{\partial \tilde{\sigma}_1} & \cdots & \frac{\partial c_{1K}}{\partial \tilde{\sigma}_1} \\ \frac{\partial c_{21}}{\partial \tilde{\sigma}_2} & \frac{\partial c_{22}}{\partial \tilde{\sigma}_2} & \cdots & \frac{\partial c_{2K}}{\partial \tilde{\sigma}_2} \\ \vdots & & & \vdots \\ \frac{\partial c_{K1}}{\partial \tilde{\sigma}_K} & \cdots & \cdots & \frac{\partial c_{KK}}{\partial \tilde{\sigma}_K} \end{bmatrix}$$

The elements of λ are

$$\frac{\partial EV_n}{\partial c_{0k}} = \frac{\partial \frac{\left|\boldsymbol{A}_{\bullet n=c_0}\right|}{|\boldsymbol{A}|}}{\partial c_{0k}} = \frac{\frac{\partial \left|\boldsymbol{A}_{\bullet n=c_0}\right|}{\partial c_{0k}}}{|\boldsymbol{A}|}; k = 1,...,K \text{ since } \frac{\partial |\boldsymbol{A}|}{\partial c_{0k}} = 0 \ \forall \ k \qquad \text{A(9)}$$

so that , using Proposition A2c):

$$\lambda = \left\{ \frac{1}{|\boldsymbol{A}|} \kappa_A (i,n); i = 1,...,K \right\} = \left\{ b_{ni}; i = 1,...,K \right\}.$$

The elements of Λ are $\frac{\partial EV_n}{\partial c_{ik}} = \frac{\partial \frac{\left|\boldsymbol{A}_{\bullet n=c_0}\right|}{|\boldsymbol{A}|}}{\partial c_{ik}} = \frac{\frac{\partial \left|\boldsymbol{A}_{\bullet n=c_0}\right|}{\partial c_{ik}}|\boldsymbol{A}| - \frac{\partial |\boldsymbol{A}|}{\partial c_{ik}}\left|\boldsymbol{A}_{\bullet n=c_0}\right|}{\left(|\boldsymbol{A}|\right)^2}; i,k = 1,...,K$,

for which we need the two derivatives, $\frac{\partial \left|\boldsymbol{A}_{\bullet n=c_0}\right|}{\partial c_{ik}}$ and $\frac{\partial |\boldsymbol{A}|}{\partial c_{ik}}$.

Proposition A4 implies that $\frac{\partial |\boldsymbol{A}|}{\partial c_{ik}} = tr\left[\left(\frac{\partial \boldsymbol{A}}{\partial c_{ik}}\right)' \mathrm{K}^{\boldsymbol{A}}\right]$ where $\mathrm{K}^{\boldsymbol{A}}$ is the matrix of cofactors of $\boldsymbol{A}$. Thus, $\frac{\partial |\boldsymbol{A}|}{\partial c_{ik}} = |\boldsymbol{A}| tr\left[\left(\frac{\partial \boldsymbol{A}}{\partial c_{ik}}\right)' \left(\boldsymbol{A}^{-1}\right)'\right] = |\boldsymbol{A}| tr\left[\boldsymbol{A}^{-1} \frac{\partial \boldsymbol{A}}{\partial c_{ik}}\right]$.

Turning now to the other derivative, write $\boldsymbol{A}^* = \boldsymbol{A}_{\bullet n=c_0}$ so that the notation is slightly less cluttered. Using Proposition A4 again:

$$\frac{\partial \left|\boldsymbol{A}^*\right|}{\partial c_{ik}} = tr\left[\left(\frac{\partial \boldsymbol{A}^*}{\partial c_{ik}}\right)' \mathrm{K}^{\boldsymbol{A}^*}\right] \text{ where } \mathrm{K}^{\boldsymbol{A}^*} \text{ is the matrix of cofactors of } \boldsymbol{A}^*.$$

Clearly, $\left(\frac{\partial \boldsymbol{A}^*}{\partial c_{ik}}\right) = \left(\frac{\partial \boldsymbol{A}}{\partial c_{ik}}\right)_{\bullet n=\mathbf{0}}$, while $\mathrm{K}^{\boldsymbol{A}^*} = \left|\boldsymbol{A}^*\right| \left[\left(\boldsymbol{A}^*\right)^{-1}\right]'$. Therefore,

$$\frac{\partial\left|\boldsymbol{A}^*\right|}{\partial c_{ik}}=\left|\boldsymbol{A}^*\right|tr\left[\left[\left(\frac{\partial\boldsymbol{A}}{\partial c_{ik}}\right)_{\bullet n=\boldsymbol{0}}\right]'\left[\left(\boldsymbol{A}^*\right)^{-1}\right]'\right]=\left|\boldsymbol{A}^*\right|tr\left[\left(\boldsymbol{A}^*\right)^{-1}\left(\frac{\partial\boldsymbol{A}}{\partial c_{ik}}\right)_{\bullet n=\boldsymbol{0}}\right].$$

Putting it all together, we find that:

$$\begin{aligned}\frac{\partial EV_n}{\partial c_{ik}}&=\frac{\frac{\partial\left|\boldsymbol{A}^*\right|}{\partial c_{ik}}\left|\boldsymbol{A}\right|-\frac{\partial\left|\boldsymbol{A}\right|}{\partial c_{ik}}\left|\boldsymbol{A}^*\right|}{\left(\left|\boldsymbol{A}\right|\right)^2}\\&=\frac{\left|\boldsymbol{A}^*\right|}{\left|\boldsymbol{A}\right|}tr\left[\left(\boldsymbol{A}^*\right)^{-1}\left(\frac{\partial\boldsymbol{A}}{\partial c_{ik}}\right)_{\bullet n=\boldsymbol{0}}-\boldsymbol{A}^{-1}\frac{\partial\boldsymbol{A}}{\partial c_{ik}}\right]\end{aligned}\qquad \text{A(10)}$$

Since $\boldsymbol{A}=\boldsymbol{I}-\delta\boldsymbol{C}$, this can be rewritten as

$$\frac{\partial EV_n}{\partial c_{ik}}=-\delta\frac{\left|\boldsymbol{A}^*\right|}{\left|\boldsymbol{A}\right|}tr\left[\left(\boldsymbol{A}^*\right)^{-1}\left(\frac{\partial\boldsymbol{C}}{\partial c_{ik}}\right)_{\bullet n=\boldsymbol{0}}-\boldsymbol{A}^{-1}\frac{\partial\boldsymbol{C}}{\partial c_{ik}}\right]\qquad \text{A(11)}$$

This last expression can be used to construct Λ as $\sum_{i,k}\frac{\partial EV_n}{\partial c_{ik}}\boldsymbol{E}_{ik}$.

Thus:

$$\Lambda=\frac{\partial EV_n}{\partial\boldsymbol{C}}=-\delta\frac{\left|\boldsymbol{A}^*\right|}{\left|\boldsymbol{A}\right|}\sum_{i,k}tr\left[\left(\boldsymbol{A}^*\right)^{-1}\left(\frac{\partial\boldsymbol{C}}{\partial c_{ik}}\right)_{\bullet n=\boldsymbol{0}}-\boldsymbol{A}^{-1}\frac{\partial\boldsymbol{C}}{\partial c_{ik}}\right]\boldsymbol{E}_{ik}$$

We now have expressions for all four components of A(12). Write the left hand side as $diag\left[\lambda\psi'\right]=\left\{b_{ni}\phi\left(\tilde{\sigma}_i\right)\left[U\left(0,\tilde{\sigma}_i\right)-U\left(1,\tilde{\sigma}_i\right)\right],i=1,...,K\right\}$. Let $\boldsymbol{D}_{\boldsymbol{x}}$ represent the square matrix with vector $\boldsymbol{x}$ on the diagonal, and zeros elsewhere. Then, using A(12):

$$\begin{aligned}\boldsymbol{D}_{diag\left[\lambda\psi'\right]}&=\boldsymbol{D}_{\left\{b_{ni}\phi\left(\tilde{\sigma}_i\right),i=1,...,K\right\}}\boldsymbol{D}_{\left\{\left[U\left(0,\tilde{\sigma}_i\right)-U\left(1,\tilde{\sigma}_i\right)\right],i=1,...,K\right\}}\\\Rightarrow\boldsymbol{D}_{\left\{\left[U\left(0,\tilde{\sigma}_i\right)-U\left(1,\tilde{\sigma}_i\right)\right],i=1,...,K\right\}}&=\boldsymbol{D}^{-1}_{\left\{b_{ni}\phi\left(\tilde{\sigma}_i\right),i=1,...,K\right\}}\boldsymbol{D}_{diag\left[\lambda\psi'\right]}\\&=\boldsymbol{D}^{-1}_{\left\{b_{ni}\phi\left(\tilde{\sigma}_i\right),i=1,...,K\right\}}\boldsymbol{D}_{diag\left[-\Lambda\Psi'\right]}\end{aligned}$$

Since $\boldsymbol{D}^{-1}_{\left\{b_{ni}\phi\left(\tilde{\sigma}_i\right),i=1,...,K\right\}}$ and $\boldsymbol{D}_{diag\left[-\Lambda\Psi'\right]}$ and are diagonal matrices, the last expression implies that

$$diag\left\{\boldsymbol{D}_{\left\{\left[U\left(0,\tilde{\sigma}_i\right)-U\left(1,\tilde{\sigma}_i\right)\right],i=1,\dots,K\right\}}\right\} = diag\left\{\boldsymbol{D}^{-1}_{\left\{b_{ni}\phi\left(\tilde{\sigma}_i\right),i=1,\dots,K\right\}}\boldsymbol{D}_{diag\left[\boldsymbol{\lambda\psi}'\right]}\right\}$$

$$= diag\left\{\boldsymbol{D}^{-1}_{\left\{b_{ni}\phi\left(\tilde{\sigma}_i\right),i=1,\dots,K\right\}}\right\} \odot diag\left\{\boldsymbol{D}_{diag\left[-\boldsymbol{\Lambda\Psi}'\right]}\right\}$$

$$\left\{\left[U\left(0,\tilde{\sigma}_i\right)-U\left(1,\tilde{\sigma}_i\right)\right],i=1,\dots,K\right\}$$

$$=\left\{\frac{1}{b_{ni}\phi\left(\tilde{\sigma}_i\right)},i=1,\dots,K\right\} \odot diag\left\{\delta\frac{\left|\boldsymbol{A}^*\right|}{\left|\boldsymbol{A}\right|}\sum_{i,k} tr\left[\left(\boldsymbol{A}^*\right)^{-1}\left(\frac{\partial \boldsymbol{C}}{\partial c_{ik}}\right)_{\bullet n=\mathbf{0}} - \boldsymbol{A}^{-1}\frac{\partial \boldsymbol{C}}{\partial c_{ik}}\right]\boldsymbol{E}_{ik}\boldsymbol{\Psi}'\right\}$$

A(12)

as claimed.

□

Appendix II

Proposition 2: For each i, $U(0,\tilde{\sigma}_i)-U(1,\tilde{\sigma}_i)$ is independent of the choice of n.

Proof:

Call this quantity $Udiff^n$ when state n is used to compute the utility differences.

$$Udiff^n = \delta\left\{\frac{1}{b_{ni}}\right\}_{i=1}^{K} \odot \left|A^n\right| diag\left[\left\{\left(\left[\left(A^n\right)^{-1}\right]'\right)_{\bullet n=0} - \left(A^{-1}\right)'\right\}\Psi'\right]$$

$$= \delta\left|A^n\right|\left\{\frac{1}{b_{ni}}\right\}_{i=1}^{K} \odot \left\{\sum_{j=1}^{K}\left[\bar{b}_{ji}^n - b_{ji}\right]\Psi_{ij}\right\}_{i=1}^{K} \quad \text{where } \bar{b}_{ji}^n = \begin{cases} b_{ji}^n \text{ if } j = j_i \neq n \text{ and } j = k_i \neq n \\ 0 \text{ if } j = j_i = n \text{ or } j = k_i = n \end{cases}$$

$$= \begin{cases} \delta\left|A^n\right|\left\{\dfrac{\left(b_{j_i i}^n - b_{j_i i}\right)-\left(b_{k_i i}^n - b_{k_i i}\right)}{b_{ni}}\right\}_{i=1}^{K} & \text{if } j_i \neq n \text{ and } k_i \neq n \\ \delta\left|A^n\right|\left\{\dfrac{\left(-b_{j_i i}\right)-\left(b_{k_i i}^n - b_{k_i i}\right)}{b_{ni}}\right\}_{i=1}^{K} & \text{if } j_i = n \\ \delta\left|A^n\right|\left\{\dfrac{\left(b_{j_i i}^n - b_{j_i i}\right)-\left(-b_{k_i i}\right)}{b_{ni}}\right\}_{i=1}^{K} & \text{if } k_i = n \end{cases}$$

We wish to show that for each i $Udiff^n$ - $Udiff^m$ =0. There are 4 cases, whose proofs are similar :

I: $j_i = n, k_i = m$; II: $j_i \neq n, k_i = m$; III: $j_i = n, k_i \neq m$; IV: $j_i \neq n, k_i \neq m$.

Case I:

$$\delta\left|A^n\right|\left\{\frac{\left(-b_{j_i i}\right)-\left(b_{k_i i}^n - b_{k_i i}\right)}{b_{ni}}\right\} - \delta\left|A^m\right|\left\{\frac{\left(b_{j_i i}^m - b_{j_i i}\right)-\left(-b_{k_i i}\right)}{b_{mi}}\right\}$$

$$= \delta\left[\left|A^n\right|\left\{\frac{\left(-b_{ni}\right)-\left(b_{mi}^n - b_{mi}\right)}{b_{ni}}\right\} - \left|A^m\right|\left\{\frac{\left(b_{ni}^m - b_{ni}\right)-\left(-b_{mi}\right)}{b_{mi}}\right\}\right]$$

$$= \delta\left[\frac{-\left|A^n\right|Cf_{in} - |A|Cf_{im}^n - \left|A^n\right|Cf_{im}}{Cf_{in}} - \frac{|A|Cf_{in}^m - \left|A^m\right|Cf_{in} + \left|A^m\right|Cf_{im}}{Cf_{im}}\right]$$

$$= \frac{\delta\left(\left|A^n\right|Cf_{im} - \left|A^m\right|Cf_{in}\right)}{Cf_{in}Cf_{im}}\left[Cf_{im}\frac{|A|Cf_{im}^n}{\left|A^m\right|Cf_{in} - \left|A^n\right|Cf_{im}} + Cf_{in}\frac{|A|Cf_{in}^m}{\left|A^m\right|Cf_{in} - \left|A^n\right|Cf_{im}} - \left(Cf_{in} - Cf_{im}\right)\right]$$

$$= \frac{\delta\left(\left|A^n\right|Cf_{im} - \left|A^m\right|Cf_{in}\right)}{Cf_{in}Cf_{im}}\left[Cf_{im}(-1) + Cf_{in}(+1) - \left(Cf_{in} - Cf_{im}\right)\right] = 0$$

Case II:

$$\delta\left|A^n\right|\left\{\frac{\left(b^n_{j_i i}-b_{j_i i}\right)-\left(b^n_{k_i i}-b_{k_i i}\right)}{b_{ni}}\right\} -\delta\left|A^m\right|\left\{\frac{\left(b^m_{j_i i}-b_{j_i i}\right)-\left(-b_{k_i i}\right)}{b_{mi}}\right\}$$

$$=\delta\left[\frac{-\left|A^n\right|Cf_{ij_i}-\left|A\right|\left(Cf^n_{im}-Cf^n_{ij_i}\right)-\left|A^n\right|Cf_{im}}{Cf_{in}} -\frac{\left|A\right|Cf^m_{ij_i}-\left|A^m\right|Cf_{ij_i}+\left|A^m\right|Cf_{im}}{Cf_{im}}\right]$$

$$=\frac{\delta\left(\left|A^n\right|Cf_{im}-\left|A^m\right|Cf_{in}\right)}{Cf_{in}Cf_{im}}\left[Cf_{im}\frac{\left|A\right|Cf^n_{im}}{\left|A^m\right|Cf_{in}-\left|A^n\right|Cf_{im}}-Cf_{im}\frac{\left|A\right|Cf^n_{ij_i}}{\left|A^m\right|Cf_{in}-\left|A^n\right|Cf_{im}} +Cf_{in}\frac{\left|A\right|Cf^m_{ij_i}}{\left|A^m\right|Cf_{in}-\left|A^n\right|Cf_{im}}-\left(Cf_{ij_i}-Cf_{im}\right)\right]$$

$$=\frac{\delta\left(\left|A^n\right|Cf_{im}-\left|A^m\right|Cf_{in}\right)}{Cf_{in}Cf_{im}}\left[Cf_{im}(-1)-Cf_{im}\frac{\left|A\right|Cf^n_{ij_i}}{\left|A^m\right|Cf_{in}-\left|A^n\right|Cf_{im}}+Cf_{in}\frac{\left|A\right|Cf^m_{ij_i}}{\left|A^m\right|Cf_{in}-\left|A^n\right|Cf_{im}}-\left(Cf_{ij_i}-Cf_{im}\right)\right]$$

$$=0$$

Case III follows directly from Case II, by substituting k_i for j_i and *n* for *m*.

Case IV:

$$\delta\left|A^n\right|\left\{\frac{\left(b^n_{j_i i}-b_{j_i i}\right)-\left(b^n_{k_i i}-b_{k_i i}\right)}{b_{ni}}\right\} -\delta\left|A^m\right|\left\{\frac{\left(b^m_{j_i i}-b_{j_i i}\right)-\left(b^m_{k_i i}-b_{k_i i}\right)}{b_{mi}}\right\}$$

$$=\delta\left[\frac{-\left|A^n\right|Cf_{ij_i}-\left|A\right|\left(Cf^n_{ik_i}-Cf^n_{ij_i}\right)-\left|A^n\right|Cf_{ik_i}}{Cf_{in}} -\frac{-\left|A^m\right|Cf_{ij_i}-\left|A\right|\left(Cf^m_{ij_i}-Cf^m_{ik_i}\right)+\left|A^m\right|Cf_{ik_i}}{Cf_{im}}\right]$$

$$=\frac{\delta\left(\left|A^n\right|Cf_{im}-\left|A^m\right|Cf_{in}\right)}{Cf_{in}Cf_{im}}\left[Cf_{im}\frac{\left|A\right|Cf^n_{ik_i}}{\left|A^m\right|Cf_{in}-\left|A^n\right|Cf_{im}}-Cf_{in}\frac{\left|A\right|Cf^m_{ik_i}}{\left|A^m\right|Cf_{in}-\left|A^n\right|Cf_{im}} +Cf_{in}\frac{\left|A\right|Cf^m_{ij_i}}{\left|A^m\right|Cf_{in}-\left|A^n\right|Cf_{im}}-Cf_{im}\frac{\left|A\right|Cf^n_{ij_i}}{\left|A^m\right|Cf_{in}-\left|A^n\right|Cf_{im}}-\left(Cf_{ij_i}-Cf_{ik_i}\right)\right]$$

$$=\frac{\delta\left(\left|A^n\right|Cf_{im}-\left|A^m\right|Cf_{in}\right)}{Cf_{in}Cf_{im}}\left[Cf_{ik_i}(-1)+Cf_{ij_i}(+1)-\left(Cf_{ij_i}-Cf_{ik_i}\right)\right]$$

$$=0$$

□

Appendix III

Applying quasi-hyperbolic discounting to our model gives, for time *t*:

$$EV_{j,t+\tau} = c_{0j} + \delta \sum_{k=1}^{K} c_{jk} EV_{k,t+\tau+1} \text{ for } \tau>0 \tag{13}$$

(discounting at exponential factor δ for time periods after $t+1$); and

$$EV_{jt} = c_{0j} + \beta\delta \sum_{k=1}^{K} c_{jk} EV_{k,t+1} \text{ for } \tau=0\,, \tag{14}$$

where $\beta < 1$.[7] Viewed from today, tomorrow is discounted with factor $\beta\delta$, whereas before today, the same lapse of time was discounted at factor δ.

Writing this in matrix form, with subscripts denoting time periods, gives $\boldsymbol{EV}_{t+\tau} = \boldsymbol{c}_0 + \delta \boldsymbol{C} \boldsymbol{EV}_{t+\tau+1}$ for $\tau>0$. Viewed from time *t*, and assuming stationarity we have: $\boldsymbol{EV}_{t+1} = \left(\boldsymbol{I} - \delta\boldsymbol{C}\right)^{-1} \boldsymbol{c}_0$. Using (4), the quasihyperbolic equivalent of A(11) is:

$$\boldsymbol{EV} = \left[\boldsymbol{I} + \beta\delta\boldsymbol{C}\left(\boldsymbol{I} - \delta\boldsymbol{C}\right)^{-1}\right]\boldsymbol{c}_0 \tag{15}$$

Note that when $\beta = 1$, the expression in brackets becomes[8] $\left(\boldsymbol{I} - \delta\boldsymbol{C}\right)^{-1}$, which is the inverse of matrix ***A*** in the exponential discounting model. The argument of Section 2 follows through with $\left[\boldsymbol{I} + \beta\delta\boldsymbol{C}\left(\boldsymbol{I} - \delta\boldsymbol{C}\right)^{-1}\right]^{-1}$ replacing the matrix ***A***. For this model the equivalent of Proposition I is:

Proposition III: Let $\boldsymbol{A} = \left[\boldsymbol{I} + \beta\delta\boldsymbol{C}\left(\boldsymbol{I} - \delta\boldsymbol{C}\right)^{-1}\right]^{-1}$, $\boldsymbol{B} = \boldsymbol{A}^{-1}$ with generic element b_{ij}. Let $\boldsymbol{X}_{\bullet n=y}$ denote the matrix ***X*** with its *n*'th column replaced by the vector ***y***, and $\boldsymbol{X}_{n\bullet=y}$ denote the matrix ***X*** with its *n*'th row replaced by the vector ***y***.

---

[7] Future versions of this paper will try to avoid using the same symbol for two different things! This discount factor is not the same thing as the parameter of the beta-distribution estimated in the previous section.

[8] Proof: $\left[\boldsymbol{I} + \beta\delta\boldsymbol{C}\left(\boldsymbol{I} - \delta\boldsymbol{C}\right)^{-1}\right]\left(\boldsymbol{I} - \delta\boldsymbol{C}\right) = \left[\left(\boldsymbol{I} - \delta\boldsymbol{C}\right) + \beta\delta\boldsymbol{C}\right] = \boldsymbol{I}$ for $\beta = 1$.

Finally, let $\boldsymbol{E}_{ik}$ be the i,k'th elementary matrix, and define $\boldsymbol{\Psi} = \left\{ \frac{\partial c_{ik}}{\partial \tilde{\sigma}_i} \right\}_{i=1}^{K}$. Then for each n the optimal vector of reservation levels $\{\tilde{\sigma}_i\}_{i=1}^{K}$ is the solution to:

$$\left\{ U(0,\tilde{\sigma}_i) - U(1,\tilde{\sigma}_i) \right\}_{i=1}^{K} = \left\{ \frac{1}{b_{ni}\phi(\tilde{\sigma}_i)} \right\}_{i=1}^{K} \odot diag \left\{ \beta\delta \frac{\left|\boldsymbol{A}_{\bullet n=c_0}\right|}{\left|\boldsymbol{A}\right|} \sum_{i,k} tr\left[\left(\boldsymbol{A}_{\bullet n=c_0}\right)^{-1} \Gamma_{\bullet n=0} - \boldsymbol{A}^{-1}\Gamma \right] \boldsymbol{E}_{ik} \boldsymbol{\Psi}' \right\} \quad (16)$$

where $\Gamma = \boldsymbol{A}\left(\boldsymbol{I} - \delta\boldsymbol{C}\right)^{-1} \frac{\partial \boldsymbol{C}}{\partial c_{ik}} \left(\boldsymbol{I} - \delta\boldsymbol{C}\right)^{-1} \boldsymbol{A}$.

The argument in Appendix I between Equations A(11) and A(14) does not depend on the structure of $\boldsymbol{A}$, so it still applies in the quasihyperbolic case. Returning to Equation A(14), to compute $\frac{\partial EV_n}{\partial c_{ik}}$ we need to know $\boldsymbol{A}^{-1}$, $\left(\boldsymbol{A}^*\right)^{-1}$ and $\frac{\partial \boldsymbol{A}}{\partial c_{ik}}$. Only the last of these needs any further algebraic attention. Differentiating both sides of $\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$ and re-arranging gives $\frac{\partial \boldsymbol{A}}{\partial c_{ik}} = -\boldsymbol{A}\frac{\partial \boldsymbol{A}^{-1}}{\partial c_{ik}}\boldsymbol{A} = -\beta\delta\boldsymbol{A}\frac{\partial \boldsymbol{C}\left(\boldsymbol{I} - \delta\boldsymbol{C}\right)^{-1}}{\partial c_{ik}}\boldsymbol{A}$. This expression can be simplified. Using the product rule,

$$\frac{\partial \boldsymbol{C}\left(\boldsymbol{I} - \delta\boldsymbol{C}\right)^{-1}}{\partial c_{ik}} = \frac{\partial \boldsymbol{C}}{\partial c_{ik}}\left(\boldsymbol{I} - \delta\boldsymbol{C}\right)^{-1} + \boldsymbol{C}\frac{\partial \left(\boldsymbol{I} - \delta\boldsymbol{C}\right)^{-1}}{\partial c_{ik}}.$$

Furthermore,

$$\frac{\partial \left(\boldsymbol{I} - \delta\boldsymbol{C}\right)^{-1}}{\partial c_{ik}} = \delta\left(\boldsymbol{I} - \delta\boldsymbol{C}\right)^{-1}\frac{\partial \boldsymbol{C}}{\partial c_{ik}}\left(\boldsymbol{I} - \delta\boldsymbol{C}\right)^{-1}$$

Using footnote 8 in the text, we can thus write

$$\begin{aligned} \frac{\partial \boldsymbol{C}\left(\boldsymbol{I} - \delta\boldsymbol{C}\right)^{-1}}{\partial c_{ik}} &= \frac{\partial \boldsymbol{C}}{\partial c_{ik}}\left(\boldsymbol{I} - \delta\boldsymbol{C}\right)^{-1} + \delta\boldsymbol{C}\left(\boldsymbol{I} - \delta\boldsymbol{C}\right)^{-1}\frac{\partial \boldsymbol{C}}{\partial c_{ik}}\left(\boldsymbol{I} - \delta\boldsymbol{C}\right)^{-1} \\ &= \left(\boldsymbol{I} - \delta\boldsymbol{C}\right)^{-1}\frac{\partial \boldsymbol{C}}{\partial c_{ik}}\left(\boldsymbol{I} - \delta\boldsymbol{C}\right)^{-1} \end{aligned}$$

so that

$$\frac{\partial A}{\partial c_{ik}} = -\boldsymbol{A}\frac{\partial \boldsymbol{A}^{-1}}{\partial c_{ik}}\boldsymbol{A}$$

$$= -\beta\delta \boldsymbol{A}\left(\boldsymbol{I}-\delta\boldsymbol{C}\right)^{-1}\frac{\partial \boldsymbol{C}}{\partial c_{ik}}\left(\boldsymbol{I}-\delta\boldsymbol{C}\right)^{-1}\boldsymbol{A}$$

$$\equiv -\beta\delta\Gamma$$

We have:

$$\frac{\partial EV_n}{\partial c_{ik}} = -\beta\delta\frac{\left|\boldsymbol{A}^*\right|}{\left|\boldsymbol{A}\right|}tr\left[\left(\boldsymbol{A}^*\right)^{-1}\Gamma_{\bullet n=\mathbf{0}} - \boldsymbol{A}^{-1}\Gamma\right]$$

and following the same logic as for the exponential case:

$$\Lambda = -\beta\delta\frac{\left|\boldsymbol{A}^*\right|}{\left|\boldsymbol{A}\right|}\sum_{i,k} tr\left[\left(\boldsymbol{A}^*\right)^{-1}\Gamma_{\bullet n=\mathbf{0}} - \boldsymbol{A}^{-1}\Gamma\right]\boldsymbol{E}_{ik}$$

so that:

$$\left\{U\left(0,\tilde{\sigma}_i\right)-U\left(1,\tilde{\sigma}_i\right)\right\}_{i=1}^{K} = \left\{\frac{1}{b_{ni}\phi\left(\tilde{\sigma}_i\right)}\right\}_{i=1}^{K}$$
$$\odot\, diag\left\{\beta\delta\frac{\left|\boldsymbol{A}_{\bullet n=c_0}\right|}{\left|\boldsymbol{A}\right|}\sum_{i,k} tr\left[\left(\boldsymbol{A}^*\right)^{-1}\Gamma_{\bullet n=\mathbf{0}} - \boldsymbol{A}^{-1}\Gamma\right]\boldsymbol{E}_{ik}\Psi'\right\}.$$

□