C Technical Appendix

C.1 A Jump-Only Continuous Time Model

In this section, we describe an alternative to the diffusion-based model presented in section §2, in which the DM updates her beliefs via a controlled Poisson process. For a derivation of this model, see Hébert and Woodford (2018). We informally demonstrate that, if the cost of the Poisson signal is described by a Bregman divergence, Theorem 1 continues to describe the DM's value function, even though the beliefs follow a Poisson process as opposed to a diffusion. Formally, Theorem 1 and the results in Hébert and Woodford (2018), taken together, imply this result.

We suppose that the DM's beliefs follows the stochastic process

$$dq_t = -\psi_t y_t dt + y_t dJ_t,$$

where dJ_t is a Poisson process with intensity ψ_t (controlled by the DM), and y_t is direction beliefs jump (also controlled by the DM). There is a trivial restriction to ensure beliefs stay in the simplex: $y_t + q_t \in \mathscr{P}(X)$ (let $Y(q_t)$ denote the set of y_t for which this holds). There is also a non-trivial restriction,

$$\psi_t D^*(q_t + y_t || q_t) \leq \chi,$$

where D^* is a divergence , convex in its first argument, and χ is a positive constant that indexes the tightness of the constraint.

We will assume that D^* satisfies, for all sets of signals S, all $\pi \in \mathscr{P}(S)$, and $q, q', \{q_s\}_{s\in S} \in \mathscr{P}(X)$ such that $\sum_{s\in S} \pi_s q_s = q'$,

$$D^*(q'||q) + \sum_{s\in S} \pi_s D^*(q_s||q') \geq \sum_{s\in S} \pi_s D^*(q_s||q).$$

Note that a Bregman divergence (as defined in equation (11)) satisfies this condition with equality. In Hébert and Woodford (2018), we prove that this condition leads to immediate stopping after jumps in the dynamic problem.

The remainder of the model is essentially identical to the one described in section §2.

The DM maximizes her expected payoff, subject to the aforementioned constraints:

$$V(q_t) = \sup_{\{y_s \in Y(q_s), \psi_s \geq 0\}, \tau \geq t} E_t[\hat{u}(q_\tau) - \kappa(\tau - t)].$$

Anywhere the value function is differentiable and the DM does not choose to stop, the Hamilton-Jacobi-Bellman (HJB) equation associated with this problem is

$$\sup_{y_t \in Y(q_t), \psi_t \geq 0} \psi_t (V(q_t + y_t) - V(q_t) - V_q(q_t) y_t) dt = \kappa dt,$$

subject to $\psi_t D^*(q_t + y_t || q_t) \leq \chi$.

It immediately follows, by $\kappa > 0$, that $\psi_t^* > 0$ and the constraint must bind, and thus

$$V(q_t + y_t^*) - V(q_t) - V_q(q_t) y_t^* = \theta D^*(q_t + y_t^* || q_t),$$

where $\theta = \chi^{-1} \kappa$. Optimality requires that

$$V(q_t + y_t^*) - V_q(q_t) y_t^* - \theta D_H(q_t + y_t^* || q_t) \geq V(q_t + y') - V_q(q_t) y' - \theta D^*(q_t + y' || q_t)$$

for all $y' \in Y(q_t)$.

We now define, from the divergence D^* and the initial beliefs q_0, a Bregman divergence $D_H(\cdot || \cdot)$, from an entropy function

$$H(q) = D^*(q || q_0).$$

We will guess and verify that the value function described by Theorem 1 satisfies these equations, with this entropy function. We will assume, to keep the exposition short, that the optimal posteriors are interior and that all actions are chosen with positive probability, but neither requirement is necessary.

The envelope theorem and first-order conditions from the static problem (equation (14)) apply. By the homogeneity of degree one of the H function,

$$q_a^T \cdot (u_a - \kappa - \theta H_q(q_a) + \theta H_q(q_0)) = q_a^T (u_a - \kappa) - \theta D_H(q_a || q_0).$$

Plugging this into the definition of the static value function, $V(q_0) = q_0^T \kappa$. Therefore, using the envelope theorem and the above expression,

$$q_a^T \cdot u_a - (q_a - q_0)^T \cdot V_q(q_0) - \theta D_H(q_a||q_0) - V(q_0) = 0.$$

Thus, if $V(q_a) = q_a^T \cdot u_{a,}$, this expression is

$$V(q_a) - V(q_0) - (q_a - q_0)^T V_q(q_0) - \theta D_H(q_a||q_0) = 0,$$

and q_a is a maximizer of this expression. We appeal to the "Locally Invariant Posteriors" property shown by Caplin et al. (2018b): q_a as a prior is a convex combination of the posteriors chosen from q_0, and therefore the same set of posteriors will be chosen with q_a as a prior, and hence it must be the case that $V(q_a) = q_a^T \cdot u_a$, as required.

By the definition of D_H,

$$V(q_a) - V(q_0) - (q_a - q_0)^T V_q(q_0) = \theta D^*(q_a||q_0),$$

and hence the first-order condition in the dynamic problem is satisfied. For any $q' \in \mathscr{P}(X)$,

$$V(q') - V(q_0) - (q' - q_0)^T V_q(q_0) \leq \theta D_H(q'||q_0).$$

Consequently,

$$\begin{aligned} \varepsilon(V(q') - V(q_0) - (q' - q_0)^T V_q(q_0)) + (1-\varepsilon)(V(q_0 - \varepsilon(q' - q_0)) - V(q_0) + \varepsilon(q' - q_0)^T V_q(q_0) \\ \leq \varepsilon\theta D_H(q'||q_0) + (1-\varepsilon)\theta D_H(V(q_0 - \varepsilon(q' - q_0)||q_0) \\ \leq \varepsilon\theta D^*(q'||q_0) + (1-\varepsilon)\theta D^*(V(q_0 - \varepsilon(q' - q_0)||q_0). \end{aligned}$$

Dividing by ε and taking limits,

$$V(q') - V(q_0) - (q' - q_0)^T V_q(q_0) \leq D^*(q'||q_0),$$

and hence optimality is satisfied.

Therefore, for any $y = q_a - q_0$, the static value function solves the HJB equation. Formalizing this proof would require dealing with boundaries, and verification. Both of these issues are technical but relatively straightforward in this context.

C.2 Convergence to the Continuous State Model

For each of a sequence of values for the integer M, we assume a neighborhood structure of the kind discussed in section 4.2 with $M+1$ states. The set of states is ordered, $X^M = \{0, 1, \ldots, M\}$, and each pair of adjacent states forms a neighborhood, $X_i = \{i, i+1\}$, for all $i \in \{0, 1, \ldots, M-1\}$. We will also assume that there is an $M+1$st neighborhood containing all of the states. Note that M indexes both the number of states and the number of neighborhoods. We consider the limit as $M \to \infty$.

To study this limit, we need to define how the prior beliefs, q_M, and the magnitude of the information costs vary with M. For the initial beliefs, we shall assume that there is a differentiable probability density function $q : [0,1] \to \mathbb{R}^+$, with full support on the unit interval and with a derivative that is Lipschitz continuous. Using this function, we define, for any $i \in X^M$,

$$e_i^T q_M = \int_{\frac{i}{M+1}}^{\frac{i+1}{M+1}} q(x)dx.$$

That is, for each value of M, the prior q_M is assumed to be a discrete approximation to the p.d.f. $q(x)$, which becomes increasingly accurate as $M \to \infty$.

For our neighborhood structures, we assume that that the constants associated with the cost of each neighborhood, c_j, are equal to M^2 for all $j < M$, and M^{-1} for $j = M$. In this particular example, the scaling ensures that the DM is neither able to determine the state with certainty, nor prevented from gathering any useful information, even as M is made arbitrarily large; moreover, the scaling ensures that the neighborhood containing all states plays no role in the limiting behavior, so that in the limit all information costs are local. We also scale the entire cost function by a constant, $\theta > 0$.

We also need to define the set of actions, and the utility from those actions. We will assume the set of actions, A, remains fixed as N grows, and define the utility from a particular action, in a particular state, as

$$e_i^T u_{a,M} = \frac{\int_{\frac{i}{M+1}}^{\frac{i+1}{M+1}} q(x)u_a(x)dx}{e_i^T q_M}.$$

Here, the utility $u_a : [0,1] \to \mathbb{R}$ is a bounded measurable function for each action $a \in A$.[32]

[32]Note that we do not require the payoff resulting from an action to be a continuous function of x at all points, though it will be continuous almost everywhere. This allows for the possibility that a DM's payoffs change discontinuously when the state x crosses some threshold, as in some of our applications.

In other words, as M grows large, the prior converges to $q(x)$ and the utilities converge to the functions $u_a(x)$.

We consider only the case of neighborhood cost functions with $\rho = 1$. Under these assumptions, the static model of Theorem 1 can be written as

$$V_N(q_M;M) = \max_{\pi_M \in \mathscr{P}(A), \{q_{a,M} \in \mathscr{P}(X^M)\}_{a\in A}} \sum_{a\in A} \pi_M(a)(u_{a,M}^T \cdot q_{a,M}) - \theta \sum_{a\in A} \pi_M(a) D_N(q_{a,M}||q_M;M), \tag{23}$$

subject to the constraint that

$$\sum_{a\in A} \pi_N(a) q_{a,M} = q_M.$$

Here D_N denotes the divergence associated with the neighborhood-based cost function introduced above, specialized to the particular neighborhood structure of this section and $\rho = 1$:

$$D_N(q_{a,M}||q_M;M) = M^2(H_N(q_{a,M};1,M) - H_N(q_M;1,M)) + M^{-1}(H^S(q_M;M) - H^S(q_{a,M};M)),$$

where H_N is defined by equation (18) in the main text and H^S is Shannon's entropy.

The following theorem shows that the solution to this problem, both in terms of the value function and the optimal policies, converges to the solution of a static rational inattention problem with a continuous state space.

Theorem 2. *Consider the sequence of finite-state-space static rational inattention problems (23), with progressively larger state spaces indexed by the natural numbers M. There exists a sub-sequence of integers $n \in \mathbb{N}$ for which the solutions to the sub-sequence of problems converge, in the sense that, for some $\pi^* \in \mathscr{P}(A)$ and $\{q_a^* \in \mathscr{P}([0,1])\}_{a\in A}$,*

i) $\lim_{n\to\infty} V_N(q_n;n) = V_N(q)$;

ii) $\lim_{n\to\infty} \pi_n^* = \pi^*$; *and*

iii) for all $a \in A$ and all $x \in [0,1]$, $\lim_{n\to\infty} \sum_{i=0}^{\lfloor xn \rfloor} e_i^T q_{a,n}^* = \int_0^x q_a^*(y)dy$.

Moreover, the limiting value function $V_N(q)$ is the value function for the following continuous-

state-space static rational inattention problem:

$$V_N(q) = \sup_{\pi \in \mathscr{P}(A), \{q_a \in \mathscr{P}_{LipG}([0,1])\}_{a \in A}} \sum_{a \in A} \pi(a) \int_{supp(q)} u_a(x) q_a(x) dx$$
$$- \frac{\theta}{4} \sum_{a \in A} \{\pi(a) \int_0^1 \frac{(q_a'(x))^2}{q_a(x)} dx\} + \frac{\theta}{4} \int_0^1 \frac{(q'(x))^2}{q(x)} dx,$$

subject to the constraint that, for all $x \in [0,1]$,

$$\sum_{a \in A} \pi(a) q_a(x) = q(x), \tag{24}$$

and where $\mathscr{P}_{LipG}([0,1])$ *denotes the set of differentiable probability density functions with full support on* $[0,1]$*, whose derivatives are Lipschitz-continuous. Furthermore, the limiting action probabilities* $\pi^*(a)$ *and posteriors* q_a^* *are the optimal policies for this continuous-state-space problem.*

Proof. See the technical appendix, section C.6. □

This theorem demonstrates that the value function, choice probabilities, and posterior beliefs of the discrete state problem converge to the value function, choice probabilities, and posterior beliefs associated with a continuous state problem. The continuous state problem uses a particular cost function, the expected value of the Fisher information $I^{Fisher}(x;p)$, defined locally for each element of the continuum of possible states x, with the expectation taken with respect to the prior over possible states. The posterior beliefs in the continuous state problem, $q_a(x)$, are required to be differentiable, with a Lipschitz-continuous derivative, on their support. This is a result; the limiting posterior beliefs of the discrete state problem will have these properties. This restriction also ensures that the Fisher information is finite, so that the optimization associated with the continuous state problem is well-behaved.

The static rational inattention problem for the limiting case of a continuous state space can be given an alternative, equivalent formulation, in which the objects of choice are the conditional probabilities of taking different actions in the different possible states, rather than the posteriors associated with different actions. This is essentially the continuous state analog of Lemma 3.

Lemma 4. *Consider the alternative continuous-state-space static rational inattention problem:*

$$\bar{V}_N(q) \;=\; \sup_{p\in\mathscr{P}_{LipG}(A)} \int_0^1 q(x)\sum_{a\in A} p_a(x)u_a(x)dx - \frac{\theta}{4}\int_0^1 q(x)\, I^{Fisher}(x;p)dx,$$

where $\mathscr{P}_{LipG}(A)$ is the set of mappings $p:[0,1]\to\mathscr{P}(A)$ such that for each action a, the function $p_a(x)$[33] is a differentiable function of x with a Lipschitz-continuous derivative, and for any information structure $p\in\mathscr{P}_{LipG}(A)$, the Fisher information *at state $x\in X$ is defined as*

$$I^{Fisher}(x;p) \;\equiv\; \sum_{a\in A}\frac{(p_a'(x))^2}{p_a(x)}.$$

This problem is equivalent to the one defined in Theorem 2, in the sense that the information structure p^* that solves this problem defines action probabilities and posteriors

$$\pi^*(a) \;=\; \int_0^1 q(x)p_a^*(x), \qquad q_a^*(x) \;=\; \frac{q(x)p_a^*(x)}{\pi^*(a)} \tag{25}$$

that solve the problem in Theorem 2, and conversely, the action probabilities and posteriors $\{\pi^*(a),q_a^*\}$ that solve the problem stated in the theorem define state-contingent action probabilities

$$p_a^*(x) \;=\; \frac{\pi^*(a)q_a^*(x)}{q(x)} \tag{26}$$

that solve the problem stated here. Moreover, the maximum achievable value is the same for both problems: $\bar{V}_N(q)=V_N(q)$.

Proof. See the appendix, section C.7. □

C.3 Security Design and Acceptance with Certainty

In this section, we discuss the optimal security design application, and consider the possibility that the seller designs the security to induce the buyer to accept with probability one. In other words, the buyer's "consideration set" in his rational inattention problem consists only of L, instead of both L and R. As mentioned in the text, we have chosen the parameters of our numerical example to ensure that, for all of the cost functions, the seller is better off

[33] Here for any $x\in[0,1]$, we use the notation $p_a(x)$ to indicate the probability of action a implied by the probability distribution $p(x)\in\mathscr{P}(A)$.

inducing information acquisition ($\pi_L < 1$) than avoiding information acquisition ($\pi_L = 1$). Note that the $\pi_L = 0$ case is equivalent to trading a "nothing" security at zero price, and hence assuming $\pi_L > 0$ is without loss of generality.

Consider the buyer's problem,

$$V(q;s,K) = \max_{\pi_L \in [0,1], q_L, q_R \in \mathscr{P}(X)} \pi_L q_L^T (s - K\iota) - \theta \pi_L D_H(q_L || q) - \theta (1 - \pi_L) D_H(q_R || q),$$

subject to the constraint that $\pi_L q_L + (1 - \pi_L) q_R = q$. Rewrite the choice variables as $\hat{q}_L = \pi_L q_L$ and $\hat{q}_R = (1 - \pi_L) q_R$, and use the homogeneity of the H function, so that the problem is

$$V(q;s,K) = \max_{\hat{q}_L, \hat{q}_R \in \mathbb{R}_+^{|X|}} \hat{q}_L^T (s - K\iota) - \theta D_H(\hat{q}_L || q) - \theta D_H(\hat{q}_R || q),$$

subject to $\hat{q}_L + \hat{q}_R = q$. Observe that the objective is concave and the constraints linear, so it suffices to consider local perturbations.

Suppose that it is optimal to set $\pi_L = 1$, implying $\hat{q}_L = q$. Consider a perturbation to $\hat{q}_L = q - \varepsilon q_R$, $\hat{q}_R = \varepsilon q_R$, for any arbitrary $q_R \in \mathscr{P}(X)$. For such a perturbation to reduce utility, we must have

$$-\varepsilon q_R^T (s - K\iota) - \theta D_H(q - \varepsilon q_R || q) - \theta \varepsilon D_H(q_R || q) \leq 0.$$

Taking the limit as $\varepsilon \to 0^+$, we must have, for all q_R, and hence for the minimizer,

$$\min_{q_R \in \mathscr{P}(X)} q_R^T (s - K\iota) + \theta D_H(q_R || q) \geq 0.$$

If this condition is satisfied, it is at least weakly optimal for the buyer to choose $\pi_L = 1$ and gather no information. Consequently, the Lagrangian version of the optimal security design problem, subject to the constraint of inducing no information acquisition, is

$$\max_{s \in \mathbb{R}_+^{|X|}, K \geq 0} \min_{\lambda \geq 0, q_R \in \mathscr{P}(X), \omega \in \mathbb{R}_+^{|X|}} q^T (K\iota - \beta s) + \lambda (q_R^T (s - K\iota) + \theta D_H(q_R || q)) + \omega^T (v - s),$$

where λ is the multiplier on the no-information-gathering constraint and ω is the multiplier on the upper-bound of the limited liability requirement.

Defining $\tilde{q}_R = \lambda q_R$, the dual of this problem is

$$\min_{\tilde{q}_R \in \mathbb{R}_+^{|X|}, \omega \in \mathbb{R}_+^{|X|}} \max_{s \in \mathbb{R}_+^{|X|}, K \geq 0} q^T (K\iota - \beta s) + \tilde{q}_R^T (s - K\iota) + \theta D_H(\tilde{q}_R || q) + \omega^T (v - s),$$

which can be understood as

$$\min_{\tilde{q}_R \in \mathbb{R}_+^{|X|}, \omega \in \mathbb{R}_+^{|X|}} \theta D_H(\tilde{q}_R || q) + \omega^T v,$$

subject to

$$\tilde{q}_R - \beta q - \omega \leq 0,$$

$$1 - q_R^T \iota \leq 0.$$

The multipliers of this convex minimization problem are the optimal security design and price. After solving the problem for $\tilde{q}_R$ and ω, we can use the first-order condition to recover the security design:

$$s - K\iota = H_q(q) - H_q(\tilde{q}_R).$$

We use the convention that in the lowest state, the asset value is zero ($e_0^T v = 0$), and therefore $e_0^T s = 0$, and hence

$$e_x^T s = (e_x - e_0)^T (H_q(q) - H_q(\tilde{q}_R)).$$

To implement the problem with the additional requirement of monotonicity for the security design, write the monotonicity requirement as $Ms \gg 0$, where M is an $|X| - 1 \times |X|$ matrix. The dual problem is

$$\min_{\tilde{q}_R \in \mathbb{R}_+^{|X|}, \omega \in \mathbb{R}_+^{|X|}, \rho \in \mathbb{R}_+^{|X|}} \theta D_H(\tilde{q}_R || q) + \omega^T v,$$

subject to

$$\tilde{q}_R - \beta q - \omega + M^T \rho \leq 0,$$

$$1 - q_R^T \iota \leq 0.$$

As mentioned above, under our parameters it is not optimal for the seller to avoid information acquisition. For completeness, we present the optimal securities that avoid information acquisition below. Note the shapes of these securities are very similar to their optimal counterparts, although the level is often quite difference.

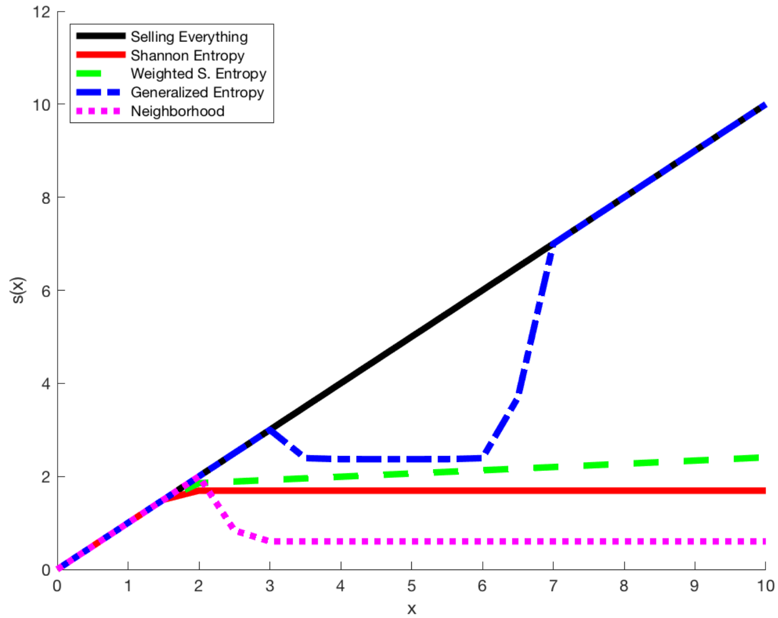

Figure 5: Optimal Security Designs that Avoid Info. Acquisition by Entropy Function

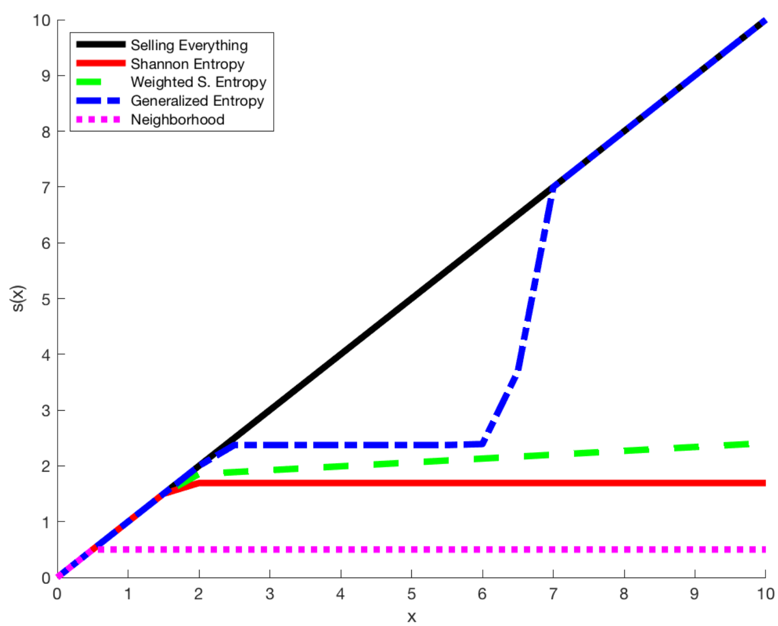

Figure 6: Optimal Monotone Security Designs that Avoid Info Acquisition by Entropy Function

C.4 The Linear-Quadratic-Gaussian Tracking Problem

Here we solve the problem in the calculus of variations stated in Section 5.4. We begin by noting that the objective (21) that we wish to minimize is of the form

$$\int_X q(x) \int_A F(a, p_a(x), p_a'(x); x)\, da\, dx,$$

where for each pair (x,a), the function

$$F(a, f, g; x) \;\equiv\; f \cdot (a-x)^2 + \frac{\theta}{4}\frac{g^2}{f}$$

is a convex function of the arguments (f,g) everywhere on its domain (the half-plane on which $f > 0$). This can be seen from the fact that (for any fixed values of (x,a)) $F(f,g)$ is equal to f times a convex function of g/f.

Given the convexity of the objective, the first-order conditions are both necessary and sufficient for an optimum. The relevant first-order conditions are furthermore the same as those for minimization of the Lagrangian

$$\int_X q(x) \int_A L(a, p_a(x), p_a'(x); x)\, da\, dx,$$

where

$$L(a, f, g; x) \;=\; F(a, f, g; x) + \varphi(x) f. \tag{27}$$

Here $\varphi(x)$ is the Lagrange multiplier associated with the constraint

$$\int_A p_a(x) da \;=\; 1 \tag{28}$$

for each $x \in X$, as is required in order for $p_a(x)$ to be a probability density function.

For given Lagrange multipliers, the problem of minimizing the Lagrangian can further be expressed as a separate minimization problem for each possible action a. Then if we can find a function $\varphi(x)$ and a function $p_a(x)$ for each $a \in A$, with $p_a(x) > 0$ for all x, such that (i) for each $a \in A$, the function $p_a(x)$ minimizes

$$\int_X q(x) L(a, p_a(x), p_a'(x); x)\, dx, \tag{29}$$

and (ii) condition (28) holds for all $x \in X$, then we will have derived an optimal information structure.

For the problem of choosing a function $p_a(x)$ to minimize (29), the first-order conditions are given by the Euler-Lagrange equations

$$q(x)\,\frac{\partial L}{\partial f}(a, p_a(x), p_a'(x); x) \;=\; \frac{d}{dx}\left[q(x)\,\frac{\partial L}{\partial g}(a, p_a(x), p_a'(x); x)\right],$$

or equivalently,

$$\frac{\partial L}{\partial f}(a, p_a(x), p_a'(x); x) \;=\; \frac{\partial L}{\partial g}(a, p_a(x), p_a'(x); x)\cdot\frac{d}{dx}[\log q(x)] + \frac{d}{dx}\left[\frac{\partial L}{\partial g}(a, p_a(x), p_a'(x); x)\right].$$

In the case of the objective function (27), we have

$$\frac{\partial L}{\partial f} \;=\; (a-x)^2 - \frac{\theta}{4}\,(v_a'(x))^2 + \varphi(x),$$

$$\frac{\partial L}{\partial g} \;=\; \frac{\theta}{2}\,v_a'(x),$$

where $v_a(x) \equiv \log p_a(x)$. Under our assumption of a Gaussian prior, we also have

$$\frac{d}{dx}[\log q(x)] \;=\; \frac{\mu - x}{\sigma^2}.$$

Substituting these expressions, the Euler-Lagrange equations take the form

$$(a-x)^2 + \varphi(x) - \frac{\theta}{4}(v_a'(x))^2 \;=\; \frac{\theta}{2}\frac{\mu - x}{\sigma^2}v_a'(x) + \frac{\theta}{2}v_a''(x)$$

for all x and a.

In the case that $\theta < 4\sigma^4$, these equations have a solution given by

$$v_a'(x) \;=\; \lambda\,[a - \beta x - (1-\beta)\mu], \tag{30}$$

$$\varphi(x) \;=\; [\beta x + (1-\beta)\mu][2 - (\beta x + (1-\beta)\mu)] - x^2 - 2\beta(1-\beta)\sigma^2,$$

where

$$\lambda \equiv \frac{2}{\theta^{1/2}} > 0, \qquad \beta \equiv 1 - \frac{\theta^{1/2}}{2\sigma^2}, \tag{31}$$

which implies (given the bound on θ) that $0 < \beta < 1$. Equation (30) is further observed to correspond to the density function $p_a(x)$ for a Gaussian distribution with mean

$$\mathrm{E}[a|x] \;=\; \beta x + (1-\beta)\mu \tag{32}$$

and variance

$$\mathrm{var}[a|x] \;=\; \frac{\beta}{\lambda} \;=\; \sigma^2\beta(1-\beta) > 0. \tag{33}$$

This solution for the distribution of a conditional on x further corresponds to a noisy representation of the state, $s = x + \varepsilon$, where the "observation error" ε is normally distributed, with mean zero and a variance ν^2, and independent of the value of x; and an estimate a of the state given by the expectation of x conditional on the noisy representation:

$$a \;=\; \mathrm{E}[x|s] \;=\; \beta s + (1-\beta)\mu. \tag{34}$$

(This is of course the estimate that minimizes the mean squared error, under the constraint that the estimate must be a function of s.)

The second equality in (34) holds if and only the variance of the observation error satisfies

$$\frac{\nu^2}{\sigma^2} \;=\; \beta^{-1} - 1 > 0. \tag{35}$$

The decision rule (34) then implies that the distribution of a conditional on x will be Gaussian, with the moments (32)–(33).

Comparison of (35) with (31) indicates that the optimal degree of noise in the representation s is given by

$$\frac{\nu^2}{\sigma^2} \;=\; [2\sigma^2\theta^{-1/2} - 1]^{-1},$$

as stated in the text. This is an increasing function of the information cost parameter θ, that approaches zero (the limiting case of perfectly accurate representation, and hence perfectly accurate estimation of the state) as θ approaches zero, and becomes unboundedly large (the limiting case of a completely uninformative information structure) as θ approaches the upper bound $4\sigma^4$ from below.

In the case that $\theta \geq 4\sigma^4$, instead, there is no solution to the Euler-Lagrange equations, and we can show that there is no interior solution to the optimization problem. Instead, as stated in the text, it is optimal to choose a completely uninformative information structure, and to choose the estimate $a = \mu$ at all times. This is because in this case, one can show that any information structure and estimation rule implies that

$$V \equiv \mathrm{E}[(a-x)^2] + \frac{\theta}{4}\mathrm{E}[I(x)] \geq \mathrm{E}[(x-\mu)^2] = \sigma^2,$$

with the lower bound achieved only in the case that $a = \mu$ with probability 1.

To prove this, we begin by observing that the Cramér-Rao bound for a biased estimator[34] implies that

$$\mathrm{E}[(a-x)^2|x] \geq \frac{(\bar{a}'(x))^2}{I(x)} + (\bar{a}(x)-x)^2,$$

where $\bar{a}(x) \equiv \mathrm{E}[a|x]$, and $I(x)$ is the Fisher information. Thus

$$\begin{aligned}
\mathrm{E}[(a-x)^2|x] + \frac{\theta}{4}I(x) &\geq \frac{(\bar{a}'(x))^2}{I(x)} + \frac{\theta}{4}I(x) + (\bar{a}(x)-x)^2 \\
&\geq \min_I\{\frac{(\bar{a}'(x))^2}{I} + \frac{\theta}{4}I\} + (\bar{a}(x)-x)^2 \\
&= \theta^{1/2}\,|\bar{a}'(x)| + (\bar{a}(x)-x)^2 \\
&\geq 2\sigma^2\,|\bar{a}'(x)| + (\bar{a}(x)-x)^2 \\
&\geq 2\sigma^2\,\bar{a}'(x) + (\bar{a}(x)-x)^2,
\end{aligned}$$

where the next-to-last inequality follows from the assumption that $\theta \geq 4\sigma^4$. Taking the expected value under the prior $q(x)$, it then follows that

$$V \geq \int_{-\infty}^{\infty} q(x)\,[2\sigma^2\,\bar{a}'(x) + (\bar{a}(x)-x)^2]\,dx. \qquad (36)$$

We wish to obtain a lower bound for the integral on the right-hand side of (36). To do this, we solve for the function $\bar{a}(x)$ that minimizes this integral, using the calculus of variations. Once again, we note that the integrand is a convex function of $\bar{a}$ and $\bar{a}'$, so that the first-order conditions are both necessary and sufficient for a minimum. The first-order conditions are given by the Euler-Lagrange equations

$$2q(x)(\bar{a}(x)-x) = 2\sigma^2 q'(x),$$

which have a unique solution $\bar{a}(x) = \mu$ for all x.

[34] See Cover and Thomas (2006), p. 396.

Substituting this solution into the integral (36), we obtain the tighter lower bound

$$V \geq \int_{\infty}^{\infty} q(x)\,(x-\mu)^2\,dx = \sigma^2. \tag{37}$$

But this lower bound is achievable by choosing $a = \mu$ with probability 1, regardless of the value of x (the optimal estimate in the case of a perfectly uninformative information structure). Hence a perfectly uninformative information structure is optimal for all $\theta \geq 4\sigma^4$.

This solution is not only *one* way of achieving the lower bound, it is the only way. It follows from the reasoning used to derive the lower bound for V that the lower bound can be achieved only if each of the weak inequalities holds as an equality. But the bound in (37) is equal to the bound in (36) only if $\bar{a}(x) = \mu$ almost surely; thus optimality requires this. And the restriction that $\mathrm{E}[a|x] = \mu$ for a set of x with full measure implies that we must have

$$\mathrm{E}[(a-x)^2|x] = (x-\mu)^2 + \mathrm{var}[a|x].$$

This in turn implies that

$$\mathrm{E}[(a-x)^2] = \mathrm{E}[(x-\mu)^2] + \mathrm{E}[\mathrm{var}[a|x]] = \sigma^2 + \mathrm{E}[\mathrm{var}[a|x]].$$

Hence the lower bound can be achieved only if $\mathrm{E}[\mathrm{var}[a|x]] = 0$.

Given that the variance is necessarily non-negative, this requires that $\mathrm{var}[a|x] = 0$ almost surely. This together with the requirement that $\mathrm{E}[a|x] = \mu$ almost surely implies that $a = \mu$ almost surely. Hence optimality requires that $a = \mu$ with probability 1, whenever $\theta \geq 4\sigma^4$.

C.5 Additional Definition and Lemmas

Definition 1. Let X^M be a sequence of state spaces, as described in section 5.3. A sequence of policies $\{p_M \in \mathscr{P}(X^M)\}_{M\in\mathbb{N}}$ satisfies the "convergence condition" if:

i) The sequence satisfies, for some constants $c_H > c_L > 0$, all M, and all $i \in X^M$,

$$\frac{c_H}{M+1} \geq e_i^T p_M \geq \frac{c_L}{M+1}.$$

ii) The sequence satisfies, for some constant $K_1 > 0$, all M, and all $i \in X^M \setminus \{0, M\}$,

$$M^3 |\frac{1}{2}(e_{i+1}^T + e_{i-1}^T - 2e_i^T) p_M| \leq K_1,$$

and

$$M^2 |\frac{1}{2}(e_M^T - e_{M-1}^T) p_M| \leq K_1$$

and

$$M^2 |\frac{1}{2}(e_1^T - e_0^T) p_M| \leq K_1.$$

Definition 2. Let $\{p_M \in \mathscr{P}(X^M)\}_{M \in \mathbb{N}}$ be a sequence of probability distributions over the state spaces associated with Theorem 2. The interpolating functions $\{\hat{p}_M \in \mathscr{P}([0,1])\}_{M \in \mathbb{N}}$ are, for $x \in [\frac{1}{2(M+1)}, 1 - \frac{1}{2(M+1)})$,

$$\hat{p}_M(x) = (M+1)((M+1)x + \frac{1}{2} - \lfloor (M+1)x + \frac{1}{2} \rfloor) e^T_{\lfloor (M+1)x + \frac{1}{2} \rfloor} p_M + \\ + (M+1)(\frac{1}{2} - (M+1)x + \lfloor (M+1)x + \frac{1}{2} \rfloor) e^T_{\lfloor (M+1)x + \frac{1}{2} \rfloor - 1} p_M,$$

and, for $x \in [0, \frac{1}{2(M+1)})$,

$$\hat{p}_M(x) = (M+1) e_0^T q_M,$$

and. for $x \in [1 - \frac{1}{2(M+1)}, 1]$,

$$\hat{p}_M(x) = (M+1) e_M^T q_M.$$

Lemma 5. *Given a function $p \in \mathscr{P}([0,1])$, define the sequence $\{p_M \in \mathscr{P}(X^M)\}_{M \in \mathbb{N}}$,*

$$e_i^T p_M = \int_{\frac{i}{M+1}}^{\frac{i+1}{M+1}} p(x) dx,$$

where X^M is the state space described in section 5.3. If the function p is strictly greater than zero for all $x \in [0,1]$, differentiable, and its derivative is Lipschitz continuous, then the sequence $\{p_M \in \mathscr{P}(X^M)\}_{N \in \mathbb{N}}$ satisfies the convergence condition, and satisfies, for some constant $K > 0$, all M, and all $i \in X^N \setminus \{0, M\}$,

$$M^2 |\ln(\frac{1}{2}(e_{i+1}^T + e_i^T) q_M) + \ln(\frac{1}{2}(e_{i-1}^T + e_i^T) q_M) - 2\ln(e_i^T q_M)| \leq K,$$

and

$$M|\ln(\frac{1}{2}(e_1^T + e_0^T)q_M) - \ln(e_0^T q_M))| < K$$

and

$$M|\ln(\frac{1}{2}(e_M^T + e_{M-1}^T)q_M) - \ln(e_M^T q_M))| < K.$$

Proof. See the technical appendix, C.8. □

Lemma 6. *Let $\{p_M \in \mathscr{P}(X^M)\}_{M\in\mathbb{N}}$ be a sequence of probability distributions over the state spaces associated with Theorem 2. If the sequence $\{p_M \in \mathscr{P}(X^M)\}_{M\in\mathbb{N}}$ satisfies the convergence condition (Definition 1), then there exists a sub-sequence, whose elements we denote by n, such that:*

i) The interpolating functions (2) $\hat{p}_n(x)$ converge point-wise to a differentiable function $p(x) \in \mathscr{P}([0,1])$, whose derivative is Lipschitz-continuous, with $p(x) > 0$ for all $x \in [0,1]$,

ii) the following sum converges:

$$\lim_{n\to\infty} n^2 \sum_{i\in X^n\setminus\{n\}} \{g(e_i^T p_n) + g(e_{i+1}^T p_n) - 2g(\frac{1}{2}(e_i^T + e_{i+1}^T)p_n)\} = \frac{1}{4}\int_0^1 \frac{(p'(x))^2}{p(x)}dx,$$

where $g(x) = x\ln(x)$,

iii) for all $a \in A$, $\lim_{n\to\infty} u_{a,n}^T p_n = \int_0^1 u_a(x)p(x)dx$,

iv) and, if the sequence $\{p_M \in \mathscr{P}(X^M)\}_{M\in\mathbb{N}}$ is constructed from some function $\tilde{p}(x)$, as in Lemma 5, then $p(x) = \tilde{p}(x)$ for all $x \in [0,1]$.

Proof. See the technical appendix, section C.9. □

Lemma 7. *Let $\pi_M(a) \in \mathscr{P}(A)$ and $\{q_{a,M} \in \mathscr{P}(X^M)\}_{a\in A}$ denote optimal policies in the discrete state setting described in section 5.3. For each $a \in A$, the sequence $\{q_{a,N}\}$ satisfies the convergence condition (Definition 1).*

Proof. See the technical appendix, section C.10. □

C.6 Proof of Theorem 2

By the boundedness of $\mathscr{P}(A)$, there exists a convergent sub-sequence of the optimal policy $\pi_n(a)$, which we also denote by n. Define

$$\pi(a) = \lim_{n\to\infty} \pi_n(a).$$

By Lemma 7, for all $a \in A$, each sequence of optimal policies $\{q_{a,n}\}$ satisfies the convergence condition (Definition 1). Therefore, by Lemma 6, each sequence of interpolating functions (2), $\{\hat{q}_{a,n}(x)\}$, has a convergent sub-sequence that converges to a differentiable function $q_a(x)$, whose derivative is Lipschitz continuous. We can construct a sub-sequence in which $\pi_n(a)$ and all $\{\hat{q}_{a,n}(x)\}$ converge by iteratively applying this argument. Pass to this subsequence.

We can write the discrete value function, using Lemma 2, and defining $g(x) = x \ln x$, as

$$\begin{aligned}
V_N(q_n;n) = & \max_{\{p_{x,n}\in\mathscr{P}(A)\}_{i\in X}} \sum_{a\in A} e_a^T p_n Diag(q) u_n e_a \\
& - \theta n^2 \sum_{a\in A} (e_a^T p_n q_n) \sum_{i=0}^{n-1} [g(\frac{e_i^T q_{a,n}}{\bar{q}_{i,a,n}}) + g(\frac{e_{i+1}^T q_{a,n}}{\bar{q}_{i,a,n}})] \\
& + \theta n^2 \sum_{i=0}^{n-1} [g(\frac{e_i^T q_N}{\bar{q}_{i,a,N}}) + g(\frac{e_{i+1}^T q_N}{\bar{q}_{i,a,N}})] \\
& - \theta n^{-1} \sum_{i=0}^{n-1} (e_i^T q_n) D_{KL}(p_n e_i || p_n q_n).
\end{aligned}$$

We can re-arrange this to

$$\begin{aligned}
V_N(q_n;n) = & \max_{\{p_{x,n}\in\mathscr{P}(A)\}_{i\in X}} \sum_{a\in A} e_a^T p_n Diag(q) u_n e_a \\
& - \theta n^2 \sum_{a\in A} (e_a^T pq) \sum_{i=0}^{n-1} [g(e_i^T q_{a,n}) + g(e_{i+1}^T q_{a,n}) - 2g(\frac{1}{2}(e_i^T + e_{i+1}^T) q_{a,n})] \\
& + \theta n^2 \sum_{i=0}^{N-1} [g(e_i^T q_n) + g(e_{i+1}^T q_n) - 2g(\frac{1}{2}(e_i^T + e_{i+1}^T) q_n)] \\
& - \theta n^{-1} \sum_{i=0}^{N-1} (e_i^T q_N) D_{KL}(p_{i,n} || p_n q_n).
\end{aligned}$$

By Lemma 6 and the boundedness of the KL divergence,

$$\lim_{n\to\infty} V_N(q_n;n) = \sum_{a\in A} \pi(a) \int_0^1 u_a(x) q_a(x) dx \\ - \frac{\theta}{4} \sum_{a\in A} \{\pi(a) \int_0^1 \frac{(q_a'(x))^2}{q_a(x)} dx\} + \frac{\theta}{4} \int_0^1 \frac{(q'(x))^2}{q(x)} dx.$$

Suppose that $\pi(a)$ and the $q_a(x)$ functions do not maximize this expression (subject to the constraints stated in Theorem 2). Let $\pi^*(a)$ and $q_a^*(x)$ be maximizers. Define, for all n,

$$\tilde{\pi}_n(a) = \pi^*(a),$$

$$e_i^T \tilde{q}_{a,n} = \int_{\frac{i}{n+1}}^{\frac{i+1}{n+1}} q_a^*(x) dx.$$

Note that, by construction, $\tilde{q}_{a,n} \in \mathscr{P}(X^n)$ and $\sum_{a\in A} \tilde{\pi}_N(a) \tilde{q}_{a,n} = q_n$. That is, the constraints of the discrete-state problem are satisfied for all n. Denote the value function under these policies as $\tilde{V}_N(q_n;n)$.

Because of the constraints stated in Theorem 2, each q_a^* satisfies the conditions of Lemma 5, and therefore the sequence $\tilde{q}_{a,n}$ satisfies the convergence condition for all $a \in A$. It follows by Lemma 6 that this sequence of policies delivers, in the limit, the value function $V_N(q)$. If this function is strictly larger than $\lim_{n\to\infty} V_N(q_n;n)$, there must exist some $\bar{n}$ such that

$$\tilde{V}_N(q_{\bar{n}};\bar{n}) > V_N(q_{\bar{n}};\bar{n}),$$

contradicting optimality. Therefore, the functions $q_a(x)$ and $\pi(a)$ are maximizers.

It remains to show that

$$\lim_{n\to\infty} \sum_{i=0}^{\lfloor xn \rfloor} e_i^T q_{a,n} = \int_0^x q_a(y) dy.$$

Note that

$$e_i^T q_{a,n} = (n+1) \int_{\frac{i}{n+1}}^{\frac{i+1}{n+1}} \hat{q}_{a,n}(\frac{2i+1}{2(n+1)}) dy,$$

where $\hat{q}_{a,n}$ is the function defined in Lemma 6. Therefore, the sum is equal to

$$\sum_{i=0}^{\lfloor xn \rfloor} e_i^T q_{a,n} = \int_0^{\frac{\lfloor xn \rfloor + 1}{n+1}} \hat{q}_{a,n}\left(\frac{\lfloor (n+1)y + \frac{1}{2} \rfloor + \frac{1}{2}}{(n+1)}\right) dy.$$

By the boundedness of $\hat{q}_{a,n}$ (which follows from the convergence condition) and the dominated convergence theorem,

$$\lim_{n\to\infty} \int_0^{\frac{\lfloor xn \rfloor + 1}{n+1}} \hat{q}\left(\frac{\lfloor (n+1)y + \frac{1}{2} \rfloor + \frac{1}{2}}{(n+1)}\right) dy = \int_0^x q_a(y) dy,$$

as required.

C.7 Proof of Lemma 4

We begin by observing that any information structure $p \in \mathscr{P}_{LipG}(A)$ defines unconditional action frequencies $\pi \in \mathscr{P}(A)$ and posteriors $q_a \in \mathscr{P}_{LipG}([0,1])$ satisfying (24), using definitions (25). And conversely, any unconditional action frequencies and posteriors satisfying (24) define an information structure, using definitions (26). Hence the set of candidate structures is the same in both problems, and the problems are equivalent if the two objective functions are equivalent as well. It is also easily seen that in each problem, the first term of the objective function is the expected value of the DM's reward $u(x,a)$, integrating over the joint distribution for (x,a). Hence it remains only to establish that the remaining terms of the objective function are equivalent as well.

Consider any information structure $p \in \mathscr{P}_{LipG}(A)$ and the corresponding unconditional action frequencies and posteriors, and let x be any point at which $q(x) > 0$, and at which $p_a(x)$ is twice differentiable for all a (and as a consequence, $q_a(x)$ is twice differentiable for all a as well). (We note that, given the Lipschitz continuity of the first derivatives, the set of x for which this is true must be of full measure.) Then the fact that $\sum_{a\in A} p_a(x) = 1$ for all x implies that

$$\sum_{a\in A} p_a''(x) = 0, \tag{38}$$

and similarly, constraint (24) implies that

$$\sum_{a\in A} \pi(a) q_a''(x) = q''(x). \tag{39}$$

At any such point, the definition of the Fisher information implies that

$$
\begin{aligned}
I^{Fisher}(x) &\equiv \sum_{a\in A} \frac{(p_a'(x))^2}{p_a(x)} \\
&= \sum_a p_a''(x) - \sum_{a\in A} p_a(x) \frac{\partial^2 \log p_a(x)}{\partial x^2} \\
&= -\frac{\pi(a) q_a(x)}{q(x)} \frac{\partial^2}{\partial x^2}[\log \pi(a) + \log q_a(x) - \log q(x)] \\
&= \frac{1}{q(x)} \left[\sum_{a\in A} \pi(a) \frac{(q_a'(x))^2}{q_a(x)} - \sum_{a\in A} \pi(a) q_a''(x) - \frac{(q'(x))^2}{q(x)} + q''(x) \right] \\
&= \frac{1}{q(x)} \left[\sum_{a\in A} \pi(a) \frac{(q_a'(x))^2}{q_a(x)} - \frac{(q'(x))^2}{q(x)} \right].
\end{aligned}
$$

Here the first line is the definition of the Fisher information (given in the lemma), and the second line follows from twice differentiating the function $\log p_a(x)$ with respect to x. In the third line, the first term from the second line vanishes because of (38); the remaining term from the second line is rewritten using (26). The fourth line follows from the third line by twice differentiating each of the terms inside the square brackets with respect to x. The fifth line then follows from (39).

Since this result holds for a set of x of full measure, we obtain expression

$$
\int_0^1 q(x) I^{Fisher}(x) dx = \sum_{a\in A} \pi(a) \int_0^1 \frac{(q_a'(x))^2}{q_a(x)} dx - \int_0^1 \frac{(q'(x))^2}{q(x)} dx
$$

for the mean Fisher information. This shows that the information-cost terms in both objective functions are equivalent, and hence the two problems are equivalent, and have equivalent solutions.

C.8 Proof of Lemma 5

Proof. The function p is strictly greater than zero, and continuous, and therefore attains a maximum and minimum on $[0,1]$, which we denote with c_H and c_L, respectively. By construction,

$$
e_i^T p_M \geq \frac{c_L}{M+1}
$$

and likewise for c_H, satisfying the bounds.

For all $i \in X^M \setminus \{M\}$,

$$\begin{aligned}(e_{i+1}^T - e_i^T)p_M &= \int_{\frac{i}{M+1}}^{\frac{i+1}{M+1}} (p(x+\frac{1}{M+1}) - p(x))dx \\ &= \int_{\frac{i}{M+1}}^{\frac{i+1}{M+1}} \int_0^{\frac{1}{M+1}} p'(x+y)dydx\end{aligned}$$

and therefore, letting K_2 be the maximum of the absolute value of p' on $[0,1]$ (which exists by the continuity of p'), we have

$$|(e_{i+1}^T - e_i^T)p_M| \leq \frac{1}{(M+1)^2}K_2, \tag{40}$$

satisfying the convergence condition for the endpoints.

For all $i \in X^M \setminus \{0, M\}$,

$$\begin{aligned}(e_{i+1}^T + e_{i-1}^T - 2e_i^T)p_M &= \int_{\frac{i}{M+1}}^{\frac{i+1}{M+1}} (p(x+\frac{1}{M+1}) + p(x-\frac{1}{M+1}) - 2p(x))dx \\ &= \int_{\frac{i}{M+1}}^{\frac{i+1}{M+1}} \int_0^{\frac{1}{M+1}} (p'(x+y) - p'(x-y))dydx.\end{aligned}$$

Let K_3 denote the Lipschitz constant associated with p'. It follows that

$$|(e_{i+1}^T + e_{i-1}^T - 2e_i^T)p_M| \leq \frac{2K_3}{(M+1)^3}.$$

Therefore, the convergence condition is satisfied for $K_1 = \max(\frac{1}{2}K_2, K_3)$.

By the concavity of the log function, and the inequality $\ln(x) \leq x - 1$,

$$\begin{aligned}\ln(\frac{\frac{1}{2}(e_{i+1}^T + e_i^T)p_M}{e_i^T p_M}) + \ln(\frac{\frac{1}{2}(e_{i-1}^T + e_i^T)p_M}{e_i^T p_M}) &\leq 2\ln(\frac{\frac{1}{4}(e_{i+1}^T + e_{i-1} + 2e_i^T)p_M}{e_i^T p_M}) \\ &\leq \frac{\frac{1}{2}(e_{i+1}^T + e_{i-1} - 2e_i^T)p_M}{e_i^T p_M}.\end{aligned}$$

Therefore, by the convergence condition we have established,

$$\ln(\frac{\frac{1}{2}(e_{i+1}^T + e_i^T)p_M}{e_i^T p_M}) + \ln(\frac{\frac{1}{2}(e_{i-1}^T + e_i^T)p_M}{e_i^T p_M}) \leq \frac{(M+1)K_1}{M^3 c_L} \leq \frac{2K_1}{M^2 c_L}.$$

By the inequality $-\ln(\frac{1}{x}) \leq x-1$,

$$\ln(\frac{\frac{1}{2}(e_{i+1}^T + e_i^T)p_M}{e_i^T p_M}) + \ln(\frac{\frac{1}{2}(e_{i-1}^T + e_i^T)p_M}{e_i^T p_M}) \geq \frac{\frac{1}{2}(e_{i+1}^T - e_i^T)p_M}{\frac{1}{2}(e_{i+1}^T + e_i^T)p_M} + \frac{\frac{1}{2}(e_{i-1}^T - e_i^T)p_M}{\frac{1}{2}(e_{i-1}^T + e_i^T)p_M}.$$

We can rewrite this as

$$\ln(\frac{\frac{1}{2}(e_{i+1}^T + e_i^T)p_M}{e_i^T p_M}) + \ln(\frac{\frac{1}{2}(e_{i-1}^T + e_i^T)p_M}{e_i^T p_M}) \geq \\ (\frac{\frac{1}{2}(e_{i+1}^T + e_{i-1}^T - 2e_i^T)p_M}{\frac{1}{2}(e_{i+1}^T + e_i^T)p_M} + \frac{\frac{1}{2}(e_{i-1}^T - e_i^T)p_M}{\frac{1}{2}(e_{i+1}^T + e_i^T)p_M}(\frac{\frac{1}{2}(e_{i+1}^T + e_i^T)p_M}{\frac{1}{2}(e_{i-1}^T + e_i^T)p_M} - 1)).$$

By the bounds above,

$$\frac{\frac{1}{2}(e_{i+1}^T + e_{i-1}^T - 2e_i^T)p_M}{\frac{1}{2}(e_{i+1}^T + e_i^T)p_M} \geq -\frac{2K_1}{M^2 c_L}$$

and, using equation (40),

$$\begin{aligned} \frac{\frac{1}{2}(e_{i-1}^T - e_i^T)p_M}{\frac{1}{2}(e_{i+1}^T + e_i^T)p_M}(\frac{\frac{1}{2}(e_{i+1}^T + e_i^T)p_M}{\frac{1}{2}(e_{i-1}^T + e_i^T)p_M} - 1) &= \frac{\frac{1}{2}(e_{i-1}^T - e_i^T)p_M}{\frac{1}{2}(e_{i+1}^T + e_i^T)p_M}(\frac{\frac{1}{2}(e_{i+1}^T - e_{i-1}^T)p_M}{\frac{1}{2}(e_{i-1}^T + e_i^T)p_M}) \\ &\geq -\frac{M^2}{c_L^2}\frac{1}{(M+1)^4}(K_2)^2 \\ &\geq -(\frac{K_2}{2Mc_L})^2. \end{aligned}$$

Therefore,

$$M^2|\ln(\frac{\frac{1}{2}(e_{i+1}^T + e_i^T)p_M}{e_i^T p_M}) + \ln(\frac{\frac{1}{2}(e_{i-1}^T + e_i^T)p_M}{e_i^T p_M})| \leq \frac{2K_1}{c_L} + (\frac{K_2}{2c_L})^2.$$

For the end-points,

$$\frac{\frac{1}{2}(e_1^T - e_0^T)q_M}{\frac{1}{2}(e_1^T + e_0^T)q_M} \leq \ln(\frac{\frac{1}{2}(e_1^T + e_0^T)q_M}{e_0^T q_M}) \leq \frac{\frac{1}{2}(e_1^T - e_0^T)q_M}{e_0^T q_M}$$

and therefore

$$|\ln(\frac{\frac{1}{2}(e_1^T + e_0^T)q_M}{e_0^T q_M})| \leq \frac{K_2}{Mc_L}.$$

A similar property holds for the other endpoint, and therefore the claim holds for $K =$

$\max(\frac{K_2}{c_L}, \frac{2K_1}{c_L} + (\frac{K_2}{2c_L})^2)$. □

C.9 Proof of Lemma 6

Proof. We begin by noting that the functions $\hat{p}_M(x)$ are absolutely continuous. Almost everywhere in $[\frac{1}{2(M+1)}, 1 - \frac{1}{2(M+1)}]$,

$$\hat{p}'_M(x) = (M+1)^2(e^T_{\lfloor (M+1)x+\frac{1}{2}\rfloor} - e^T_{\lfloor (M+1)x+\frac{1}{2}\rfloor - 1})p_M,$$

and outside this region, $\hat{p}'_M(x) = 0$. Let $\tilde{p}'_M(x)$ denote the right-continuous Lebesgue-integrable function on $[0,1]$ such that

$$\hat{p}_M(x) = \hat{p}_M(0) + \int_0^x \tilde{p}'_M(y)dy,$$

which is equal to $\hat{p}'_M(x)$ anywhere the latter exists.

The total variation of $\tilde{p}'_M(x)$ is equal to

$$TV(\tilde{p}'_M) = \sum_{i=1}^{M-1} (M+1)^2|(e^T_{i+1} + e^T_{i-1} - 2e^T_i)p_M)| + \\ + (M+1)^2|(e^T_M - e^T_{M-1})p_M| + (M+1)^2|(e^T_1 - e^T_0)p_M|.$$

By the convergence condition,

$$TV(\tilde{p}'_M) \leq \frac{(M+1)^3}{M^3} 2K_1,$$

and therefore the sequence of functions $\tilde{p}'_M(x)$ has uniformly bounded variation.

For any $1 - \frac{1}{2(M+1)} > x > y \geq \frac{1}{2(M+1)}$, the quantity

$$|\tilde{p}'_M(x) - \tilde{p}'_M(y)| = (M+1)^2 |\sum_{i=\lfloor (M+1)y+\frac{1}{2}\rfloor}^{\lfloor (M+1)x+\frac{1}{2}\rfloor} (e^T_{i+1} + e^T_{i-1} - 2e^T_i)p_M| \\ \leq \frac{(M+1)^2((M+1)(x-y)+2)}{M^3} 2K_1.$$

At the end points, for all $x \in [0, \frac{1}{2(M+1)})$,

$$|\tilde{p}'_M(\frac{1}{2(M+1)}) - \tilde{p}'_M(x)| \leq \frac{2K_1}{M+1},$$

and for all $x \in [1 - \frac{1}{2(M+1)}, 1]$,

$$|\tilde{p}'_M(x) - \lim_{y \uparrow 1 - \frac{1}{2(M+1)}} \tilde{p}'_M(y)| \leq \frac{2K_1}{M+1}.$$

By $\tilde{p}'_M(0) = 0$, we have, for all $x \in [0, 1]$,

$$|\tilde{p}'_M(x)| \leq (\frac{(M+1)^2((M+1)(1 - \frac{1}{2(M+1)}) + 2)}{M^3} + \frac{1}{M+1})2K_1,$$

proving that $\tilde{p}'_M(x)$ is bounded uniformly in M for all $x \in [0, 1]$.

Therefore Helly's selection theorem applies. That is, there exists a sub-sequence, which we denote by n, such that $\tilde{p}'_n(x)$ converges point-wise to some $p'(x)$. Moreover, by the point-wise convergence of $\tilde{p}'_M$ to p', for all $x > y$,

$$|p'(x) - p'(y)| \leq 2K_1(x - y),$$

meaning that p' is Lipschitz-continuous. By the fact that $p'(0) = 0$, this implies that $|p'(x)| \leq 2K_1$ for all $x \in [0, 1]$.

By the convergence condition, $c_L \leq \hat{p}_N(0) \leq c_H$. Therefore, there exists a convergent sub-sequence. We now use n to denote the sub-sequence for which $\lim_{n\to\infty} \hat{p}_n(0) = p(0)$ and for which $\tilde{p}'_n(x)$ converges point-wise to $p'(x)$. By the dominated convergence theorem, for all $x \in [0, 1]$,

$$\lim_{n\to\infty} \hat{p}_n(x) = \lim_{n\to\infty} \{\hat{p}_n(0) + \int_0^x \tilde{p}'_n(y)dy\} = p(0) + \int_0^x p'(y)dy.$$

Define the function $p(x) = p(0) + \int_0^x p'(y)dy$ for all $x \in [0, 1]$. By the convergence conditions, this function is bounded, $0 < c_L \leq p(x) \leq c_H$, by construction it is differentiable, and its derivative is Lipschitz continuous. Moreover,

$$\int_0^1 p(x)dx = 1,$$

and therefore $p \in \mathscr{P}([0,1])$.

Next, consider the limiting cost function. We have, using the function $g(x) = x \ln x$ and Taylor-expanding,

$$g(y) = g(x) + g'(x)(y-x) + \frac{1}{2} g''(cy + (1-c)x)(y-x)^2$$

for some $c \in (0,1)$. Therefore,

$$\begin{aligned} g(e_i^T p_M) + g(e_{i+1}^T p_M) - 2g(\frac{1}{2}(e_i^T + e_{i+1}^T) p_M) = \\ \frac{1}{8} g''(c_1 e_i^T p_M + (1-c_1)\frac{1}{2}(e_i^T + e_{i+1}^T) p_M)((e_{i+1}^T - e_i^T) p_M)^2 \\ + \frac{1}{8} g''(c_2 e_i^T p_M + (1-c_2)\frac{1}{2}(e_i^T + e_{i+1}^T) p_M)((e_{i+1}^T - e_i^T) p_M)^2 \end{aligned}$$

for constants $c_1, c_2 \in (0,1)$. Note that, by the boundedness $\hat{p}_M(x)$ from below, $e_i^T p_M \geq (M+1)^{-1} c_L$ for all $i \in X^M$. It follows that

$$g''(c_1 e_i^T p_M + (1-c_1)\frac{1}{2}(e_i^T + e_{i+1}^T) p_M) = \frac{1}{c_1 e_i^T p_M + (1-c_1)\frac{1}{2}(e_i^T + e_{i+1}^T) p_M} \leq (M+1) c_L^{-1}.$$

Therefore,

$$0 \leq g(e_i^T p_M) + g(e_{i+1}^T p_M) - 2g(\frac{1}{2}(e_i^T + e_{i+1}^T) p_M) \leq \frac{(M+1) c_L^{-1}}{4} ((e_{i+1}^T - e_i^T) p_M)^2.$$

By construction,

$$e_i^T p_M = \frac{1}{(M+1)} \hat{p}_M(\frac{2i+1}{2(M+1)}).$$

Therefore,

$$\begin{aligned} (M+1)(g(e_i^T p_M) + g(e_{i+1}^T p_M) - 2g(\frac{1}{2}(e_i^T + e_{i+1}^T) p_M)) = \\ g(\hat{p}_M(\frac{2i+1}{2(M+1)})) + g(\hat{p}_M(\frac{2i+3}{2(M+1)})) - 2g(\hat{p}_M(\frac{2i+2}{2(M+1)})). \end{aligned}$$

and

$$g(e_i^T p_M) + g(e_{i+1}^T p_M) - 2g(\frac{1}{2}(e_i^T + e_{i+1}^T) p_M) \leq \frac{c_L^{-1}}{4(M+1)} (\hat{p}(\frac{2i+3}{2(M+1)}) - \hat{p}(\frac{2i+1}{2(M+1)}))^2.$$

By the boundedness of $\tilde{p}'_M(x)$,

$$g(\hat{p}(\frac{2i+1}{2(M+1)})) + g(\hat{p}(\frac{2i+3}{2(M+1)})) - 2g(\hat{p}(\frac{2i+2}{2(M+1)})) \leq \frac{B}{(M+1)^2}$$

for some finite bound B.

Writing the limiting cost as an integral, and switching to the sub-sequence n defined above,

$$n^2 \sum_{i \in X^n \setminus \{n\}} \{g(e_i^T p_n) + g(e_{i+1}^T p_n) - 2g(\frac{1}{2}(e_i^T + e_{i+1}^T) p_n)\} =$$
$$\frac{n^3}{n+1} \int_0^1 \{g(\hat{p}_n(\frac{2\lfloor nx \rfloor + 1}{2(n+1)})) + g(\hat{p}_n(\frac{2\lfloor nx \rfloor + 3}{2(n+1)})) - 2g(\hat{p}_n(\frac{2\lfloor nx \rfloor + 2}{2(n+1)}))\} dx.$$

By the bound above,

$$\frac{n^3}{n+1} \int_0^1 \{g(\hat{p}_n(\frac{2\lfloor nx \rfloor + 1}{2(n+1)})) + g(\hat{p}_n(\frac{2\lfloor nx \rfloor + 3}{2(n+1)})) - 2g(\hat{p}_n(\frac{2\lfloor nx \rfloor + 2}{2(n+1)}))\} dx \leq$$
$$\frac{n^3}{(n+1)^3} \int_0^1 B dx.$$

Applying the dominated convergence theorem,

$$\lim_{n \to \infty} n^2 \sum_{i \in X^n \setminus \{n\}} \{g(e_i^T p_n) + g(e_{i+1}^T p_n) - 2g(\frac{1}{2}(e_i^T + e_{i+1}^T) p_n)\} =$$
$$\int_0^1 \lim_{n \to \infty} \frac{n^3}{n+1} \{g(\hat{p}_n(\frac{2\lfloor nx \rfloor + 1}{2(n+1)})) + g(\hat{p}_n(\frac{2\lfloor nx \rfloor + 3}{2(n+1)})) - 2g(\hat{p}_n(\frac{2\lfloor nx \rfloor + 2}{2(n+1)}))\} dx.$$

By the Taylor expansion above,

$$\lim_{n \to \infty} \frac{n^3}{n+1} \{g(\hat{p}_n(\frac{2\lfloor nx \rfloor + 1}{2(n+1)})) + g(\hat{p}_n(\frac{2\lfloor nx \rfloor + 3}{2(n+1)})) - 2g(\hat{p}_n(\frac{2\lfloor nx \rfloor + 2}{2(n+1)}))\} =$$
$$\lim_{n \to \infty} \frac{1}{8} \frac{n^3}{n+1} \{g''(\cdot) + g''(\cdot)\} (\hat{p}_n(\frac{2\lfloor nx \rfloor + 3}{2(n+1)}) - \hat{p}_n(\frac{2\lfloor nx \rfloor + 1}{2(n+1)}))^2.$$

By definition,

$$(n+1)(\hat{p}_n(\frac{2\lfloor nx \rfloor + 3}{2(n+1)}) - \hat{p}_n(\frac{2\lfloor nx \rfloor + 1}{2(n+1)})) = \tilde{p}'_n(\frac{2\lfloor nx \rfloor + 2}{2(n+1)})$$

and

$$\lim_{n\to\infty} g''(\hat{p}_n(\frac{2\lfloor nx\rfloor+2}{2(n+1)})+c_n(\hat{p}_n(\frac{2\lfloor nx\rfloor+3}{2(n+1)})-\hat{p}_n(\frac{2\lfloor nx\rfloor+2}{2(n+1)}))) = \frac{1}{p(x)},$$

with $c_n \in (0,1)$ for all n, and therefore

$$\lim_{n\to\infty} \frac{n^3}{n+1}\{g(\hat{p}_n(\frac{2\lfloor nx\rfloor+1}{2(n+1)}))+g(\hat{p}_n(\frac{2\lfloor nx\rfloor+3}{2(n+1)}))-2g(\hat{p}_n(\frac{2\lfloor nx\rfloor+2}{2(n+1)}))\} =$$

$$\lim_{n\to\infty} \frac{1}{4}\frac{(p'(x))^2}{p(x)},$$

proving the second claim.

Turning to the third claim, recall that, by definition,

$$e_i^T u_{a,M} = \frac{\int_{\frac{i}{M+1}}^{\frac{i+1}{M+1}} u_a(x)q(x)dx}{\int_{\frac{i}{M+1}}^{\frac{i+1}{M+1}} q(x)dx.}$$

We define the function, for $x \in [0,1)$, as

$$u_{a,M}(x) = e^T_{\lfloor (M+1)x\rfloor} u_{a,M},$$

and let $u_{a,M}(1) = e_M^T u_{a,M}$. We also define the function

$$\tilde{x}_M(x) = \frac{2\lfloor (M+1)x\rfloor+1}{2(M+1)}.$$

By construction, $\hat{p}_M(\tilde{x}_M(x)) = (M+1)e^T_{\lfloor (M+1)x\rfloor} p_{a,M}$ for all $x \in [0,1)$, and equals $e_M^T p_{a,M}$ for $x=1$. Therefore,

$$\begin{aligned} u_{a,M}^T p_M &= \sum_{i\in X^M} (e_i^T u_{a,M})(e_i^T p_M) \\ &= \int_0^1 \hat{p}_M(\tilde{x}_M(x)) u_{a,M}(x)dx. \end{aligned}$$

By the measurability of $u_a(x)$,

$$\lim_{M\to\infty} u_{a,M}(x) = u_a(x).$$

Therefore, by the boundedness of utilities and the dominated convergence theorem,

$$\lim_{n\to\infty} u_{a,n}^T p_n = \int_0^1 p(x)u_a(x)dx.$$

Finally, suppose that, for all M

$$e_i^T p_{a,M} = \int_{\frac{i}{M+1}}^{\frac{i+1}{M+1}} \tilde{p}(x)dx.$$

It follows that $\lim_{n\to\infty} \hat{p}_{a,n}(x) = \tilde{p}(x)$ for all $x \in [0,1]$, and therefore $\tilde{p}(x) = p(x)$. □

C.10 Proof of Lemma 7

Proof. We begin by noting that the conditions given for the function $q(x)$ satisfy the conditions of Lemma 5, and therefore the sequence q_M satisfies the convergence condition. We will use the constants c_H and c_L to refer to its bounds,

$$\frac{c_H}{M+1} \geq e_i^T q_M \geq \frac{c_L}{M+1},$$

and the constants K_1 and K to refer to the constants described by convergence condition and Lemma 5 for the sequence q_M. By the convention that $q_{a,M} = q_M$ if $\pi_M(a) = 0$, $q_{a,M}$ also satisfies the convergence condition whenever $\pi_M(a) = 0$.

The problem of size M is

$$V_N(q_M;M) = \max_{\pi_M \in \mathscr{P}(A), \{q_{a,M} \in \mathscr{P}(X^M)\}_{a\in A}} \sum_{a\in A} \pi_M(a)(u_{a,M}^T \cdot q_{a,M}) - \theta \sum_{a\in A} \pi_M(a) D_N(q_{a,M}||q_M;M)$$

subject to

$$\sum_{a\in A} \pi_M(a) q_{a,M} = q_M,$$

where

$$D_N(q_{a,M}||q_M;\rho,M) = M^2(H_N(q_{a,M};1,M) - H_N(q_M;1,M)) + M^{-1}(H^S(q_{a,M};M) - H^S(q_M;M)$$

and

$$H_N(q;1,M) = -\sum_{i=0}^{M-1} \bar{q}_i H^S(q_i).$$

Let u_M denote that $|X^M| \times |A|$ matrix whose columns are $u_{a,M}$. Using Lemma 3, we can rewrite the problem as

$$
\begin{aligned}
V_N(q_M;M) = & \max_{\{p_{i,M} \in \mathscr{P}(A)\}_{i \in X^M}} \sum_{a \in A} e_a^T p_M Diag(q) u_M e_a \\
& - \theta M^2 \sum_{i=0}^{M-1} (e_i^T q_M) D_{KL}(p_{i,M} || \frac{p_{i,M}(e_i^T q_M) + p_{i+1,M}(e_{i+1}^T q_M)}{(e_i^T + e_{i+1}^T) q_M}) \\
& - \theta M^2 \sum_{i=1}^{M} (e_i^T q_M) D_{KL}(p_{i,M} || \frac{p_{i,M}(e_i^T q_N) + p_{i-1,M}(e_{i-1}^T q_M)}{(e_i^T + e_{i-1}^T) q_M}) \\
& - \theta M^{-1} \sum_{i=0}^{M-1} (e_i^T q_M) D_{KL}(p_{i,M} || p_M q_M).
\end{aligned}
$$

The FOC for this problem is, for all $i \in [1, M-1]$ and $a \in A$ such that $e_a^T p_{i,M} > 0$,

$$
\begin{aligned}
e_i^T u_{a,M} - \theta M^2 \ln(\frac{e_a^T p_{i,M}(e_i^T + e_{i+1}^T) q_M}{e_a^T (p_{i,M}(e_i^T q_M) + p_{i+1,M}(e_{i+1}^T q_M))}) \\
-\theta M^2 \ln(\frac{e_a^T p_{i,M}(e_i^T + e_{i-1}^T) q_M}{e_a^T (p_{i,M}(e_i^T q_M) + p_{i-1,N}(e_{i-1}^T q_M))}) - \theta M^{-1} \ln(\frac{e_a^T p_{i,M}}{e_a^T p_M q_M}) - e_i^T \kappa_M = 0,
\end{aligned}
$$

where $\kappa_M \in \mathbb{R}^{M+1}$ are the multipliers (scaled by $e_i^T q_M$) on the constraints that $\sum_{a \in A} e_a^T p_{i,M} = 1$ for all $i \in X$. Defining $e_{i-1}^T q_M = e_{M+1}^T q_M = 0$, and defining $p_{-1,M}$ and $p_{M+1,M}$ in arbitrary fashion, we can recover this FOC for all $i \in X$.

Rewriting the FOC in terms of the posteriors, and again defining $e_{i-1}^T q_{a,M} = e_{M+1}^T q_{a,M} = 0$, for any a such that $\pi_M(a) > 0$,

$$
\begin{aligned}
e_i^T (u_{a,M} - \kappa_M) = & \theta M^2 \ln(\frac{(e_i^T q_{a,M})(1 + \frac{e_{i+1}^T q_M}{e_i^T q_M})}{(e_{i+1} + e_i)^T q_{a,M}}) + \theta M^2 \ln(\frac{(e_i^T q_{a,N})(1 + \frac{e_{i-1}^T q_N}{e_i^T q_N})}{(e_{i-1} + e_i)^T q_{a,N}}) \\
& + \theta M^{-1} \ln(\frac{e_a^T p_{i,M}}{e_a^T p_M q_M}) \\
= & -\theta M^2 \ln(1 + \frac{e_{i+1}^T q_{a,M}}{e_i^T q_{a,M}}) + \theta M^2 \ln(1 + \frac{e_{i+1}^T q_M}{e_i^T q_M}) - \theta M^2 \ln(1 + \frac{e_{i-1}^T q_{a,M}}{e_i^T q_{a,M}}) \\
& + \theta M^2 \ln(1 + \frac{e_{i-1}^T q_M}{e_i^T q_M}) + \theta M^{-1} \ln(\frac{e_i^T q_{a,M}}{e_i^T q_M}),
\end{aligned}
$$

which can be rewritten as

$$e_i^T(u_{a,M}-\kappa_M) = -\theta M^2(\ln(\frac{1}{2}(e_{i+1}^T+e_i^T)q_{a,M})+\ln(\frac{1}{2}(e_{i-1}^T+e_i^T)q_{a,M})-(2+M^{-3})\ln(e_i^T q_{a,M}))$$
$$+\theta M^2(\ln(\frac{1}{2}(e_{i+1}^T+e_i^T)q_M)+\ln(\frac{1}{2}(e_{i-1}^T+e_i^T)q_M)-(2+M^{-3})\ln(e_i^T q_M)). \tag{41}$$

Our analysis proceeds by analyzing this first-order condition.

We next describe a series of lemmas that use this first-order condition to establish various bounds, which will ultimately be used to establish the bounds required by the convergence condition. As part of the proof, we find it useful to consider the interpolating functions $\hat{q}_{a,M}(x)$ (2) constructed from $q_{a,M}$. We define from these interpolating functions the function

$$l_{a,N}(x) = (M+1)(\ln(\hat{q}_{a,M}(x))-\ln(\hat{q}_{a,M}(x-\frac{1}{2(M+1)})))$$

on $x \in [\frac{1}{2(M+1)},1]$, observing that, for any $i \in X^M \setminus \{0\}$,

$$l_{a,M}(\frac{2i+1}{2(M+1)}) = (M+1)\ln(\frac{(M+1)e_i^T q_{a,M}}{\frac{1}{2}(M+1)(e_i^T+e_{i-1}^T)q_{a,M}}),$$

and for any $i \in X^M \setminus \{M\}$,

$$l_{a,M}(\frac{2i+2}{2(M+1)}) = (M+1)\ln(\frac{\frac{1}{2}(M+1)(e_i^T+e_{i+1}^T)q_{a,M}}{(M+1)e_i^T q_{a,M}}).$$

□

Lemma 8. *For all $M \in \mathbb{N}$ and $i \in X^M \setminus \{0,M\}$, $e_i^T\kappa_M \le B_\kappa$ for some positive constant B_κ.*

Proof. See the technical appendix, section C.11. □

Lemma 9. *For all $M \in \mathbb{N}$ and $i \in \{0,M\}$, $|e_i^T\kappa_M| \le B_0$ for some positive constant B_0, and*

$$\ln(\frac{\frac{1}{2}(e_0^T+e_1^T)q_{a,M}}{e_0^T q_{a,M}}) \le M^{-1}B_1$$

and

$$\ln(\frac{e_M^T q_{a,M}}{\frac{1}{2}(e_M^T+e_{M-1}^T)q_{a,M}}) \ge -M^{-1}B_1$$

for some positive constant B_1.

Proof. See the technical appendix, section C.12. □

Lemma 10. *For all $M \in \mathbb{N}$ and $j \in \{2,3,\ldots,2M+1\}$, and some positive constant B_l,*

$$|l_{a,N}(\frac{j}{2(M+1)})| \leq B_l.$$

Proof. See the technical appendix, section C.13. The proof uses the previous two lemmas. □

C.10.1 Proof that $\frac{c_H}{M+1} \geq e_i^T q_{a,M} \geq \frac{c_L}{M+1}$

We next apply the above lemmas to prove that the first part of the convergence condition is satisfied. Begin by observing that there must exist some $\tilde{i}_{a,M} \in X^M$ such that $e_{\tilde{i}_{a,M}}^T q_{a,M} \geq \frac{1}{N+1}$, implying that

$$\ln((M+1)e_{\tilde{i}_{a,M}}^T q_{a,M}) \geq 0.$$

By the definition of $l_{a,M}$, for any $i \in X^M \setminus \{0\}$,

$$l_{a,M}(\frac{2i+1}{2(M+1)}) + l_{a,M}(\frac{2i}{2(M+1)}) = (M+1)\ln(\frac{(M+1)e_i^T q_{a,M}}{(M+1)e_{i-1}^T q_{a,M}}).$$

For any $i > \tilde{i}_{a,M}$, using Lemma 10,

$$\begin{aligned}
\ln((M+1)e_i^T q_{a,M}) &= \ln((M+1)e_{\tilde{i}_{a,M}}^T q_{a,M}) + \sum_{j=\tilde{i}_{a,M}+1}^{i} \ln(\frac{(M+1)e_j^T q_{a,M}}{(M+1)e_{j-1}^T q_{a,M}}) \\
&= \ln((M+1)e_{\tilde{i}_{a,M}}^T q_{a,M}) + \frac{1}{M+1} \sum_{j=\tilde{i}_{a,M}+1}^{i} l_{a,M}(\frac{2j+1}{2(M+1)}) + l_{a,N}(\frac{2j}{2(M+1)}) \\
&\geq -\frac{1}{M+1} \sum_{j=\tilde{i}_{a,M}+1}^{i} 2B_l \\
&\geq -2B_l.
\end{aligned}$$

Similarly, for any $i < \tilde{i}_{a,M}$,

$$\ln((M+1)e_{\tilde{i}_{a,M}}^T q_{a,M}) = \ln((M+1)e_i^T q_{a,M}) + \sum_{j=i+1}^{\tilde{i}_{a,M}} \ln(\frac{(N+1)e_j^T q_{a,N}}{(N+1)e_{j-1}^T q_{a,N}}).$$

Therefore, for any $i < \tilde{i}_{a,M}$,

$$\ln((M+1)e_i^T q_{a,M}) \geq -\sum_{j=i+1}^{\tilde{i}_{a,M}} \ln(\frac{(M+1)e_j^T q_{a,M}}{(M+1)e_{j-1}^T q_{a,M}}),$$

and thus, using Lemma 10, for all $i \in X^M$,

$$\ln((M+1)e_i^T q_{a,M}) \geq -2B_l.$$

Repeating this argument, there must be some $\hat{i}_{a,M}$ such that $e_{\hat{i}_{a,M}}^T q_{a,M} \leq M^{-1}$, and using the bounds on $l_{a,M}$ in similar fashion yields

$$\ln((M+1)e_i^T q_{a,M}) \leq 2B_l.$$

It follows that, for all M, $a \in A$ such that $\pi_M(a) > 0$, and $i \in X^M$,

$$\frac{exp(2B_l)}{(M+1)} \geq e_i^T q_{a,M} \geq \frac{\exp(-2B_l)}{M+1}, \tag{42}$$

demonstrating that $q_{a,N}$ satisfies the first part of the convergence condition.

C.10.2 Proof that $M^3|\frac{1}{2}(e_{i+1}^T + e_{i-1}^T - 2e_i^T)q_{a,M}| \leq K_1$

We start by proving a bound on $(M+1)^2|\frac{1}{2}(e_{i+1}^T - e_i^T)q_{a,M}|$.

Using Lemma 10, and a Taylor expansion of $\ln(1+x)$, for some $c \in (0,1)$, for any $i \in X^M \setminus \{M\}$,

$$\begin{aligned}|l_{a,M}(\frac{2i+2}{2(M+1)})| &= |(M+1)\ln(\frac{\frac{1}{2}(M+1)(e_i^T + e_{i+1}^T)q_{a,M}}{(M+1)e_i^T q_{a,M}})| \\ &= \frac{(M+1)|\frac{1}{2}(e_{i+1}^T - e_i^T)q_{a,M}|}{e_i^T q_{a,M} + \frac{c}{2}(e_{i+1}^T - e_i^T)q_{a,M}} \\ &\leq B_l,\end{aligned}$$

and therefore, by the bound on $e_i^T q_{a,M}$, for any $i \in X^M \setminus \{M\}$,

$$(M+1)^2|\frac{1}{2}(e_{i+1}^T - e_i^T)q_{a,M}| \leq B_l \exp(-2B_l). \tag{43}$$

Returning to the first-order condition, for $i \in X^N \setminus \{0,N\}$, and using the bounds on utility and on the terms involving q_M,

$$\begin{aligned} e_i^T \kappa_M \geq & -\bar{u} - \theta K + \theta M^{-1} \ln(\frac{e_i^T q_M}{e_i^T q_{a,M}}) \\ & + \theta M^2 (\ln(\frac{1}{2}(e_{i+1}^T + e_i^T) q_{a,M}) + \ln(\frac{1}{2}(e_{i-1}^T + e_i^T) q_{a,M}) - 2\ln(e_i^T q_{a,M})). \end{aligned}$$

We have

$$M^{-1} \ln(\frac{e_i^T q_M}{e_i^T q_{a,M}}) \geq M^{-1} \ln(\frac{c_L}{\exp(2B_l)}),$$

and therefore

$$e_i^T \kappa_M \geq -\bar{u} - \theta K + M^{-1} \ln(\frac{c_L}{\exp(2B_l)}) + \theta M^2 (\ln(\frac{\frac{1}{2}(e_{i+1}^T + e_i^T) q_{a,M}}{e_i^T q_{a,M}}) + \ln(\frac{\frac{1}{2}(e_{i-1}^T + e_i^T) q_{a,M}}{e_i^T q_{a,M}})).$$

Using the mean-value theorem, for some $c_1 \in (0,1)$,

$$\begin{aligned} \ln(\frac{\frac{1}{2}(e_{i+1}^T + e_i^T) q_{a,M}}{e_i^T q_{a,M}}) &= \ln(1 + \frac{\frac{1}{2}(e_{i+1}^T - e_i^T) q_{a,M}}{e_i^T q_{a,M}}) \\ &= \frac{e_i^T q_{a,M}}{e_i^T q_{a,M} + c_1 \frac{1}{2}(e_{i+1}^T - e_i^T) q_{a,M}} \frac{\frac{1}{2}(e_{i+1}^T - e_i^T) q_{a,M}}{e_i^T q_{a,M}}, \end{aligned}$$

and likewise

$$\ln(\frac{\frac{1}{2}(e_{i-1}^T + e_i^T) q_{a,M}}{e_i^T q_{a,M}}) = \frac{\frac{1}{2}(e_{i-1}^T - e_i^T) q_{a,M}}{(1 - \frac{1}{2}c_2) e_i^T q_{a,M} + \frac{1}{2} c_1 e_{i-1}^T q_{a,M}}$$

for some $c_2 \in (0,1)$. Therefore,

$$\begin{aligned} e_i^T \kappa_M \geq & -\bar{u} - \theta K + M^{-1} \ln(\frac{c_L}{\exp(2B_l)}) \\ & + \theta M^2 (\frac{\frac{1}{2}(e_{i+1}^T - e_i^T) q_{a,M}}{(1 - \frac{1}{2}c_1) e_i^T q_{a,M} + \frac{1}{2} c_1 e_{i+1}^T q_{a,M}} + \frac{\frac{1}{2}(e_{i-1}^T - e_i^T) q_{a,M}}{(1 - \frac{1}{2}c_2) e_i^T q_{a,M} + \frac{1}{2} c_2 e_{i-1}^T q_{a,M}}). \end{aligned}$$

Multiplying through,

$$
\begin{aligned}
&[(1-\frac{1}{2}c_1)e_i^T q_{a,M}+\frac{1}{2}c_1 e_{i+1}^T q_{a,M}](e_i^T \kappa_M+\bar{u}+\theta K-M^{-1}\ln(\frac{c_L}{\exp(2B_l)}))\\
&\geq \theta M^2(\frac{1}{2}(e_{i+1}^T-e_i^T)q_{a,M}+\frac{1}{2}(e_{i-1}^T-e_i^T)q_{a,M}\frac{(1-\frac{1}{2}c_1)e_i^T q_{a,M}+\frac{1}{2}c_1 e_{i+1}^T q_{a,M}}{(1-\frac{1}{2}c_2)e_i^T q_{a,M}+\frac{1}{2}c_2 e_{i-1}^T q_{a,M}}).\\
&\geq \theta M^2(\frac{1}{2}(e_{i+1}^T+e_{i-1}^T-2e_i^T)q_{a,M}+\frac{1}{2}(e_{i-1}^T-e_i^T)q_{a,M}(\frac{\frac{1}{2}c_1(e_{i+1}^T-e_i^T)q_{a,M}-\frac{1}{2}c_2(e_i^T-e_{i-1}^T)q_{a,M}}{(1-\frac{1}{2}c_2)e_i^T q_{a,M}+\frac{1}{2}c_2 e_{i-1}^T q_{a,M}})).
\end{aligned}
$$

Using equations (42) and (43),

$$
\begin{aligned}
&[(1-\frac{1}{2}c_1)e_i^T q_{a,M}+\frac{1}{2}c_1 e_{i+1}^T q_{a,M}](e_i^T \kappa_M+\bar{u}+\theta K-M^{-1}\ln(\frac{c_L}{\exp(2B_l)}))\\
&\geq \theta M^2(\frac{1}{2}(e_{i+1}^T+e_{i-1}^T-2e_i^T)q_{a,M}-\frac{B_l\exp(2B_l)}{(M+1)^2}(\frac{\frac{2B_l\exp(2B_l)}{(M+1)^2}}{\frac{\exp(-2B_l)}{M+1}}))\\
&\geq \theta M^2\frac{1}{2}(e_{i+1}^T+e_{i-1}^T-2e_i^T)q_{a,M}-\theta\frac{2B_l^2M^2\exp(6B_l)}{(M+1)^3}.
\end{aligned}
$$

Summing over a, weighted by $\pi_N(a)$, and applying Lemma 5,

$$
\begin{aligned}
(e_i^T \kappa_M+\bar{u}+\theta K-M^{-1}\ln(\frac{c_L}{\exp(2B_l)})) &\geq -\theta\frac{\frac{K_1}{M}+\frac{2B_l^2M^2\exp(6B_l)}{(M+1)^3}}{\frac{c_L}{(M+1)}}\\
&\geq -\theta c_L^{-1}(2K_1+2B_l^2\exp(6B_l)).
\end{aligned}
$$

Therefore, $|e_i^T \kappa_N|$ is bounded below by some $B_\kappa^+>0$ for all $i\in X^N$ (recalling that this was shown for $i\in\{0,N\}$ in Lemma 9 and in the other direction in Lemma 8).

It also follows, using equation (42), that

$$
\begin{aligned}
\theta M^2(M+1)\frac{1}{2}(e_{i+1}^T+e_{i-1}^T-2e_i^T)q_{a,M} \leq exp(2B_l)(B_\kappa^+ +\bar{u}+\theta K-M^{-1}\ln(\frac{c_L}{\exp(2B_l)})\\
+\theta\frac{2B_l^2M^2\exp(6B_l)}{(M+1)^2},
\end{aligned}
$$

which establishes one side of the bound on $|\frac{1}{2}(e_{i+1}^T+e_{i-1}^T-2e_i^T)q_{a,M}|$.

Rewriting the FOC (equation (41)) and using Lemma 5 and the boundedness of the utility and the bound on $|e_i^T \kappa_N|$,

$$-B_\kappa^+ - \bar{u} - \theta K - \theta M^{-1} \ln(\frac{e_i^T q_M}{e_i^T q_{a,M}})$$
$$\leq \theta M^2 (\ln(\frac{1}{2}(e_{i+1}^T + e_i^T) q_{a,M}) + \ln(\frac{1}{2}(e_{i-1}^T + e_i^T) q_{a,M}) - 2\ln(e_i^T q_{a,M})).$$

By equation (42),

$$M^{-1} \ln(\frac{e_i^T q_M}{e_i^T q_{a,M}}) \leq M^{-1} \ln(\frac{c_H}{\exp(-2B_l)}),$$

and therefore, by the concavity of the log function,

$$-B_\kappa^+ - \bar{u} - \theta K - \theta M^{-1} \ln(\frac{c_H}{\exp(-2B_l)}) \leq 2\theta M^2 \ln(\frac{\frac{1}{4}(e_{i+1}^T + e_{i-1}^T + 2e_i^T) q_{a,M}}{e_i^T q_{a,M}}).$$

By the inequality $\ln(x) \leq x - 1$,

$$-B_\kappa^+ - \bar{u} - \theta K - \theta M^{-1} \ln(\frac{c_H}{\exp(-2B_l)}) \leq 2\theta M^2 (\frac{\frac{1}{4}(e_{i+1}^T + e_{i-1}^T - 2e_i^T) q_{a,M}}{e_i^T q_{a,M}}),$$

and therefore, using the lower bound on $e_i^T q_{a,M}$ (equation (42)),

$$-B_\kappa^+ - \bar{u} - \theta K - \theta M^{-1} \ln(\frac{c_H}{\exp(-2B_l)}) \leq \theta M^2 (M+1) \frac{1}{2}(e_{i+1}^T + e_{i-1}^T - 2e_i^T) q_{a,M},$$

which proves the other side of the bound.

C.10.3 Proof that $M^2 |\frac{1}{2}(e_1^T - e_0^T) q_{a,M}| \leq K_1$

By Lemma 10,

$$-B_l \leq (M+1) \ln(\frac{\frac{1}{2}(e_0^T + e_1^T) q_{a,M}}{e_0^T q_{a,M}}) \leq B_l.$$

Using the mean-value theorem, for some $c \in (0,1)$,

$$\ln(\frac{\frac{1}{2}(e_0^T + e_1^T) q_{a,M}}{e_0^T q_{a,M}}) = \frac{\frac{1}{2}(e_1^T - e_0^T) q_{a,M}}{(1 - \frac{1}{2}c) e_0^T q_{a,M} + \frac{1}{2} c e_i^T q_{a,M}}.$$

Therefore, by equation (42),

$$\frac{\exp(2B_l)}{(M+1)^2}B_l \geq \frac{1}{2}(e_1^T - e_0^T)q_{a,M} \geq -\frac{\exp(2B_l)}{(M+1)^2}B_l,$$

proving the bound. The proof for the other endpoint is identical.

C.11 Proof of Lemma 8

First, using Lemma 5, for all $i \in X^M \setminus \{0, M\}$, observe that

$$M^2|\ln(\frac{1}{2}(e_{i+1}^T + e_i^T)q_M) + \ln(\frac{1}{2}(e_{i-1}^T + e_i^T)q_M) - 2\ln(e_i^T q_M)| \leq K.$$

Rewriting the FOC (equation (41)) and using this bound,

$$\begin{aligned} e_i^T \kappa_M \leq{} & e_i^T u_{a,M} + \theta K + \theta M^{-1} \ln(e_i^T q_M) \\ & + \theta M^2(\ln(\frac{1}{2}(e_{i+1}^T + e_i^T)q_{a,M}) + \ln(\frac{1}{2}(e_{i-1}^T + e_i^T)q_{a,M}) - (2 + M^{-3})\ln(e_i^T q_{a,M})). \end{aligned}$$

By the boundedness of the utility function, this can be rewritten as

$$e_i^T \kappa_M \leq \bar{u} + \theta K - \theta M^2(\ln(\frac{e_i^T q_{a,M}}{\frac{1}{2}(e_{i+1}^T + e_i^T)q_{a,M}}) + \ln(\frac{e_i^T q_{a,M}}{\frac{1}{2}(e_{i-1}^T + e_i^T)q_{a,M}})) - \theta M^{-1}\ln(\frac{e_i^T q_{a,M}}{e_i^T q_M}).$$

By the concavity of the log function,

$$\begin{aligned} & \ln(\frac{1}{2}(e_{i+1}^T + e_i^T)q_{a,M}) + \ln(\frac{1}{2}(e_{i-1}^T + e_i^T)q_{a,M}) + M^{-3}\ln(e_i^T q_M) \leq \\ & \qquad (2 + M^{-3})\ln(\frac{1}{2(2+M^{-3})}(e_{i+1}^T + e_{i-1}^T + 2e_i^T)q_{a,M} + \frac{M^{-3}}{2+M^{-3}}e_i^T q_M), \end{aligned}$$

It follows that

$$e_i^T \kappa_N \leq \bar{u} + \theta K + (2 + M^{-3})\theta M^2 \ln(\frac{\frac{1}{2(2+M^{-3})}(e_{i+1}^T + e_{i-1}^T + 2e_i^T)q_{a,M} + \frac{N^{-3}}{2+N^{-3}}e_i^T q_M}{e_i^T q_{a,M}}).$$

Exponentiating,

$$(e_i^T q_{a,M}) \exp(-\frac{1}{2+M^{-3}} \theta^{-1} M^{-2} (\bar{u} + \bar{\theta} K - e_i^T \kappa_M)) \leq \frac{1}{2(2+M^{-3})}(e_{i+1}^T + e_{i-1}^T + 2e_i^T) q_{a,M} + \frac{M^{-3}}{2+M^{-3}} e_i^T q_M.$$

Summing over a, weighted by $\pi_N(a)$,

$$(e_i^T q_M) \exp(-\frac{1}{2+M^{-3}} \theta^{-1} M^{-2} (\bar{u} + \bar{\theta} K - e_i^T \kappa_M)) \leq \frac{1}{2(2+M^{-3})}(e_{i+1}^T + e_{i-1}^T + 2e_i^T) q_M + \frac{M^{-3}}{2+M^{-3}} e_i^T q_M.$$

Taking logs,

$$\begin{aligned} -\frac{1}{2+M^{-3}} \theta^{-1} M^{-2} (\bar{u} + \bar{\theta} K - e_i^T \kappa_M) &\leq \ln(\frac{\frac{1}{2(2+M^{-3})}(e_{i+1}^T + e_{i-1}^T + 2e_i^T) q_M + \frac{M^{-3}}{2+M^{-3}} e_i^T q_M}{(e_i^T q_M)}) \\ &\leq \ln(1 + \frac{M^{-3}}{2+M^{-3}} + \frac{1}{2+M^{-3}} \frac{K_1 M^{-3}}{c_L M^{-1}}), \end{aligned}$$

where the last step follows by Lemma 5, recalling that c_L is the lower bound on $q(x)$. We have

$$\begin{aligned} e_i^T \kappa_N &\leq 3\theta M^2 \ln(1 + \frac{M^{-3}}{2+M^{-3}} + \frac{1}{2+M^{-3}} \frac{K_1}{c_L} M^{-2}) + \bar{u} + \bar{\theta} K \\ &\leq \bar{u} + \theta K + \frac{3\theta M^{-1}}{2+M^{-3}} + \frac{3\theta}{2+M^{-3}} \frac{K_1}{c_L} \\ &\leq \bar{u} + \theta K + \frac{3\theta}{2} + \frac{3\theta}{2} \frac{K_1}{c_L}. \end{aligned}$$

where the second step follows by the inequality $\ln(1+x) < x$ for $x > 0$.

C.12 Proof of Lemma 9

For the lower end point, the FOC (equation (41)) can be simplified to

$$e_0^T(u_{a,M}-\kappa_M) = -\theta M^2(\ln(\frac{1}{2}(e_1^T+e_0^T)q_{a,M})+\ln(\frac{1}{2})-(1+M^{-3})\ln(e_0^T q_{a,M}))$$
$$+\theta M^2(\ln(\frac{1}{2}(e_1^T+e_0^T)q_M)+\ln(\frac{1}{2})-(1+M^{-3})\ln(e_0^T q_M)).$$

Rearranging this,

$$\theta^{-1}M^{-2}e_0^T(u_{a,M}-\kappa_M)+\ln(\frac{1}{2}(e_1^T+e_0^T)q_{a,M}) =$$
$$(1+M^{-3})\ln(\frac{e_0^T q_{a,M}}{e_0^T q_M})+\ln(\frac{1}{2}(e_1^T+e_0^T)q_M).$$

Exponentiating,

$$\frac{1}{2}(e_1^T+e_0^T)q_{a,M}\exp(\theta^{-1}M^{-2}e_0^T(u_{a,M}-\kappa_M)) = (\frac{e_0^T q_{a,M}}{e_0^T q_M})^{1+M^{-3}}\frac{1}{2}(e_1^T+e_0^T)q_M.$$

By the boundedness of the utility function,

$$\frac{1}{2}(e_1^T+e_0^T)q_{a,M}\exp(\theta^{-1}M^{-2}(\bar{u}-e_0^T\kappa_M)) \geq (\frac{e_0^T q_{a,M}}{e_0^T q_M})^{1+M^{-3}}\frac{1}{2}(e_1^T+e_0^T)q_M.$$

Taking a sum over a, weighted by $\pi(a)$, and applying Jensen's inequality,

$$\frac{1}{2}(e_1^T+e_0^T)q_M\exp(\theta^{-1}M^{-2}(\bar{u}-e_0^T\kappa_M)) \geq \frac{1}{2}(e_1^T+e_0^T)q_M,$$

and therefore

$$e_0^T\kappa_M \leq \bar{u}.$$

Observing that

$$M^{-1}\ln(\frac{e_0^T q_{a,M}}{e_0^T q_M}) \leq M^{-1}\ln(\frac{M}{c_L}) \leq M^{-1}(\frac{M}{c_L}-1) \leq c_L^{-1}, \tag{44}$$

we have

$$\theta^{-1}M^{-2}e_0^T(u_{a,M}-\kappa_M)+\ln(\frac{1}{2}(e_1^T+e_0^T)q_{a,M}) \leq M^{-2}c_L^{-1}+\ln(\frac{e_0^T q_{a,M}}{e_0^T q_M})+\ln(\frac{1}{2}(e_1^T+e_0^T)q_M).$$

Exponentiating,

$$\frac{1}{2}(e_1^T+e_0^T)q_{a,M}\exp(\theta^{-1}M^{-2}(-\theta c_L^{-1}+e_0^T(u_{a,M}-\kappa_M)))\leq(\frac{e_0^T q_{a,M}}{e_0^T q_M})\frac{1}{2}(e_1^T+e_0^T)q_M$$

Using the boundedness of the utility function, then taking a sum over a, weighted by $\pi(a)$,

$$\frac{1}{2}(e_1^T+e_0^T)q_{a,M}\exp(\theta^{-1}M^{-2}(-\theta c_L^{-1}-\bar{u}-e_0^T\kappa_M))\leq\frac{1}{2}(e_1^T+e_0^T)q_M.$$

Therefore,

$$e_0^T\kappa_M\geq-\bar{u}-\theta c_L^{-1},$$

and thus

$$|e_0^T\kappa_M|\leq B_0$$

for $B_0=\bar{u}+\theta c_L^{-1}$. A similar argument applies to the other end-point ($e_M^T\kappa_M$).

Using the bound on utility and equation (44), the FOC requires that

$$\ln(\frac{\frac{1}{2}(e_1^T+e_0^T)q_{a,M}}{e_0^T q_{a,M}})\leq\theta^{-1}M^{-2}(\bar{u}+B_0+\theta c_L^{-1})+\ln(\frac{\frac{1}{2}(e_1^T+e_0^T)q_M}{e_0^T q_M}).$$

By Lemma 5, it follows that

$$\ln(\frac{\frac{1}{2}(e_1^T+e_0^T)q_{a,M}}{e_0^T q_{a,M}})\leq\theta^{-1}M^{-2}(\bar{u}+B_0+\theta c_L^{-1})+M^{-1}K,$$

and therefore the constraint with $B_1=K+\theta^{-1}(\bar{u}+B_0+\theta c_L^{-1})$ is satisfied.

Similarly, the FOC for the highest state is

$$\theta^{-1}M^{-2}e_M^T(u_{a,M}-\kappa_M)+\ln(\frac{\frac{1}{2}(e_M^T+e_{M-1}^T)q_{a,M}}{e_M^T q_{a,M}})=$$
$$(1+M^{-3})\ln(\frac{e_M^T q_{a,M}}{e_M^T q_M})+\ln(\frac{1}{2}(e_M^T+e_{M-1}^T)q_M),$$

and therefore

$$\ln(\frac{\frac{1}{2}(e_M^T+e_{M-1}^T)q_{a,M}}{e_M^T q_{a,M}})\leq\theta^{-1}M^{-2}(\bar{u}+B_0+\theta c_L^{-1})+\ln(\frac{\frac{1}{2}(e_M^T+e_{M-1}^T)q_M}{e_M^T q_M}),$$

implying that

$$\ln(\frac{\frac{1}{2}(e_M^T + e_{M-1}^T)q_{a,M}}{e_M^T q_{a,M}}) \leq \theta^{-1}M^{-2}(\bar{u} + B_0 + \theta c_L^{-1}) + M^{-1}K,$$

and therefore

$$\ln(\frac{e_M^T q_{a,M}}{\frac{1}{2}(e_M^T + e_{M-1}^T)q_{a,M}}) \geq -M^{-1}B_1.$$

C.13 Proof of Lemma 10

The first-order condition is, for any $i \in X^M \setminus \{0,M\}$ can be re-written using the function $l_{a,M}$ (and the function l_M, defined from $\hat{q}_M$ along the same lines) as

$$\begin{aligned} e_i^T(\kappa_M - u_{a,M}) + \theta M^{-1}\ln(\frac{e_i^T q_{a,M}}{e_i^T q_M}) &= \theta\frac{M^2}{(M+1)}(l_{a,M}(\frac{2i+2}{2(M+1)}) - l_{a,M}(\frac{2i+1}{2(M+1)})) \\ &\quad - \theta\frac{M^2}{(M+1)}(l_M(\frac{2i+2}{2(M+1)}) - l_M(\frac{2i+1}{2(M+1)})). \end{aligned}$$

Note that

$$\theta M^{-1}\ln(\frac{e_i^T q_{a,M}}{e_i^T q_M}) \leq \theta M^{-1}\ln(\frac{1}{c_L M^{-1}}) \leq \theta M^{-1}(\frac{M}{c_L} - 1) \leq \theta c_L^{-1}.$$

By Lemma 5 and Lemma 8 and the bound on utility,

$$\theta\frac{M^2}{(M+1)}(l_{a,M}(\frac{2i+2}{2(M+1)}) - l_{a,M}(\frac{2i+1}{2(M+1)}) \leq B_\kappa + \bar{u} + \theta K + \theta c_L^{-1}.$$

We also have, for all $i \in X^M \setminus \{M\}$

$$\begin{aligned} &\frac{M^2}{M+1}(l_{a,M}(\frac{2i+3}{2(M+1)}) - l_{a,M}(\frac{2i+2}{2(M+1)})) \\ &= M^2(\ln(\frac{(M+1)e_{i+1}^T q_{a,M}}{\frac{1}{2}(M+1)(e_{i+1}^T + e_i^T)q_{a,M}}) - \ln(\frac{\frac{1}{2}(M+1)(e_i^T + e_{i+1}^T)q_{a,M}}{(M+1)e_i^T q_{a,M}})) \\ &\leq 0, \end{aligned}$$

by the concavity of the log function. Observe also that, by Lemma 9,

$$l_{a,M}(\frac{2}{2(M+1)}) = (M+1)\ln(\frac{\frac{1}{2}(e_0^T + e_1^T)q_{a,M}}{e_0^T q_{a,M}}) \leq \frac{M+1}{M}B_1.$$

It follows that, for all $j \in \{2,3,\ldots,2M+1\}$,

$$\begin{aligned} l_{a,M}(\frac{j}{2(M+1)}) &= l_{a,M}(\frac{2}{2(M+1)}) + \sum_{k=2}^{j-1}(l_{a,M}(\frac{k+1}{2(N+1)}) - l_{a,M}(\frac{k}{2(M+1)})) \\ &\leq \theta^{-1}(B_\kappa + \bar{u} + \theta K + \theta c_L^{-1})\frac{M+1}{M^2}(j-2) + \frac{M+1}{M}B_1. \end{aligned}$$

Similarly, for all $j \in \{2,3,\ldots,2M+1\}$,

$$l_{a,M}(\frac{2M+1}{2(M+1)}) = l_{a,M}(\frac{j}{2(M+1)}) + \sum_{k=j}^{2M}(l_{a,M}(\frac{k+1}{2(M+1)}) - l_{a,M}(\frac{k}{2(M+1)})).$$

Observing that

$$-l_{a,M}(\frac{2M+1}{2(M+1)}) = -\ln(\frac{(M+1)e_M^T q_{a,M}}{\frac{1}{2}(M+1)(e_M^T + e_{M-1}^T)q_{a,M}}) \leq \frac{M+1}{M}B_1,$$

using Lemma 9,

$$-l_{a,M}(\frac{j}{2(M+1)}) \leq \theta^{-1}(B_\kappa + \bar{u} + \theta K + \theta c_L^{-1})\frac{M+1}{M^2}(2M-j+1) + \frac{M+1}{M}B_1.$$

It follows that, for all $j \in \{2,3,\ldots,2M+1\}$,

$$\begin{aligned} |l_{a,N}(\frac{j}{2(N+1)})| &\leq \theta^{-1}(B_\kappa + \bar{u} + \theta K + \theta c_L^{-1})\frac{M+1}{M^2}(2M-1) + \frac{M+1}{M}B_1 \\ &\leq 4\theta^{-1}(B_\kappa + \bar{u} + \theta K + \theta c_L^{-1}) + 2B_1. \end{aligned}$$