

Comparative statics with adjustment costs and the le Chatelier principle*

Eddie Dekel[†] John K.-H. Quah[‡] Ludvig Sinander[§]

22 February 2023

Abstract: We develop a theory of monotone comparative statics for models with adjustment costs. We show that comparative-statics conclusions may be drawn under the usual ordinal complementarity assumptions on the objective function, assuming very little about costs: only a mild monotonicity condition is required. We use this insight to prove a general le Chatelier principle: under the ordinal complementarity assumptions, if short-run adjustment is subject to a monotone cost, then the long-run response to a shock is greater than the short-run response. We extend these results to a fully dynamic model of adjustment over time: the le Chatelier principle remains valid, and under slightly stronger assumptions, optimal adjustment follows a monotone path. We apply our results to models of capital investment and of sticky prices.

Keywords: Adjustment costs, comparative statics, le Chatelier.

JEL classification numbers: C02, C7, D01, D2, D8.

*We are grateful to Gregorio Curello for close reading and comments, and to Sara Neff, Simon Quinn, Jakub Steiner, Quitzé Valenzuela-Stookey, Nathaniel Ver Steeg, Jan Žemlička and audiences at Berlin, Bonn, CERGE-EI, Oxford, the 8th TUS conference and the 5th Bergamo IO Winter Symposium for helpful comments.

[†]Department of Economics, Northwestern University and Department of Economics, Tel Aviv University. Email address: <eddiedekel@gmail.com>.

[‡]Department of Economics, Johns Hopkins University and Department of Economics, National University of Singapore. Email address: <john.quah@jhu.edu>.

[§]Department of Economics and Nuffield College, University of Oxford. Email address: <ludvig.sinander@economics.ox.ac.uk>.

1 Introduction

Adjustment costs play a major role in explaining a wide range of economic phenomena. Examples include the investment behavior of firms (e.g. Jorgenson, 1963; Hayashi, 1982; Cooper & Haltiwanger, 2006), price stickiness (e.g. Mankiw, 1985; Caplin & Spulber, 1987; Golosov & Lucas, 2007; Midrigan, 2011), aggregate consumption dynamics (e.g. Kaplan & Violante, 2014; Berger & Vavra, 2015) and housing consumption and asset pricing (Grossman & Laroque, 1990).

In this paper, we develop a theory of monotone comparative statics with adjustment costs. Our fundamental insight is that, surprisingly, very little needs to be assumed about the cost function: comparative statics requires only that *not* adjusting be cheaper than adjusting, plus the usual ordinal complementarity assumptions on the objective function. We use this insight to show that Samuelson's (1947) *le Chatelier principle* is far more general than previously claimed: it holds whenever adjustment is costly, given only minimal structure on costs. We extend our comparative-statics and le Chatelier results to a fully dynamic model of adjustment.

We apply our results to models of factor demand, capital investment, and pricing. These models are typically studied only under strong functional-form assumptions, and the cases of convex and nonconvex costs are considered separately and handled very differently. Our general results yield robust comparative statics for these standard models, dispensing with auxiliary assumptions and handling convex and nonconvex costs in a unified fashion.

The abstract setting is as follows. An agent chooses an action x from a sublattice $L \subseteq \mathbb{R}^n$. Her objective $F(x, \theta)$ depends on a parameter θ . At the initial parameter $\underline{\theta}$, the agent chose $\underline{x} \in \arg \max_{x \in L} F(x, \underline{\theta})$. The parameter now increases to $\bar{\theta} > \underline{\theta}$, and the agent may adjust her choice. Adjusting the action by $\epsilon = x - \underline{x}$ costs $C(\epsilon) \geq 0$, and the agent's new choice maximizes $G(x, \bar{\theta}) = F(x, \bar{\theta}) - C(x - \underline{x})$.

Our only assumption on the cost function C is *monotonicity*:

$$C(\epsilon_1, \dots, \epsilon_{i-1}, \epsilon'_i, \epsilon_{i+1}, \dots, \epsilon_n) \leq C(\epsilon) \quad \text{whenever } 0 \leq \epsilon'_i \leq \epsilon_i \text{ or } 0 \geq \epsilon'_i \geq \epsilon_i.$$

This means that cost falls whenever an adjustment vector ϵ is modified by

shifting one of its entries closer to zero (“no adjustment”). An additively separable cost function $C(\epsilon) = \sum_{i=1}^n C_i(\epsilon_i)$ is monotone exactly if each dimension’s cost function C_i is single-dipped and minimized at zero.

By relying on monotonicity alone, we eschew restrictive assumptions such as convexity; even quasiconvexity is not needed. Some adjustments ϵ may be infeasible, as captured by a prohibitive cost $C(\epsilon) = \infty$. In some of our results, the monotonicity assumption may be weakened to *minimal monotonicity*: cost falls whenever an adjustment vector ϵ is modified by replacing all of its positive entries with zero ($C(\epsilon \wedge 0) \leq C(\epsilon)$), and similarly for the negative entries ($C(\epsilon \vee 0) \leq C(\epsilon)$). In the additively separable case, this means that each dimension’s cost C_i is minimized at zero.

Our basic question is under what assumptions on the objective F and cost C the agent’s choice increases, in the sense that $\hat{x} \geq \underline{x}$ for some $\hat{x} \in \arg \max_{x \in L} G(x, \bar{\theta})$ (provided the argmax is not empty; such qualifiers are omitted throughout this introduction). Our fundamental result, Theorem 1, answers this question: nothing need be assumed about the cost C except minimal monotonicity, while F need only satisfy the ordinal complementarity conditions of *quasi-supermodularity* and *single-crossing differences* that feature in similar comparative-statics results absent adjustment costs (see Milgrom & Shannon, 1994). Thus costs need not even be monotone, and the objective need not satisfy any cardinal properties, such as supermodularity or increasing differences. We also give a “ \forall ” variant (Proposition 1): adding either of two mild assumptions yields the stronger conclusion that $\hat{x} \geq \underline{x}$ for *every* $\hat{x} \in \arg \max_{x \in L} G(x, \bar{\theta})$.

We use our fundamental result to re-think Samuelson’s (1947) *le Chatelier principle*, which asserts that the response to a parameter shift is greater at longer horizons. Our Theorem 2 provides that the le Chatelier principle holds whenever short-run adjustment is subject to a monotone adjustment cost C , long-run adjustment is frictionless, and the objective F satisfies the ordinal complementarity conditions. Formally, the theorem states that under these assumptions, given any long-run choice $\bar{x} \in \arg \max_{x \in L} F(x, \bar{\theta})$ satisfying $\bar{x} \geq \underline{x}$, we have $\bar{x} \geq \hat{x} \geq \underline{x}$ for some optimal short-run choice $\hat{x} \in \arg \max_{x \in L} G(x, \bar{\theta})$.¹

¹Furthermore, if \bar{x} is the largest element of $\arg \max_{x \in L} F(x, \underline{x})$, then $\bar{x} \geq \hat{x}$ for *any* short-run choice $\hat{x} \in \arg \max_{x \in L} G(x, \bar{\theta})$.

This substantially generalizes Milgrom and Roberts's (1996) le Chatelier principle, in which short-run adjustment is assumed to be impossible for some dimensions i and costless for the rest: that is, $C(\epsilon) = \sum_{i=1}^n C_i(\epsilon_i)$, where some dimensions i have $C_i(\epsilon_i) = \infty$ for all $\epsilon_i \neq 0$, and the rest have $C_i \equiv 0$. We show that our le Chatelier principle remains valid if long-run adjustment is also costly (Proposition 2).

We then extend our comparative-statics and le Chatelier theorems to a fully dynamic, forward-looking model of costly adjustment over time. The parameter θ_t evolves over time $t \in \{1, 2, 3, \dots\}$, and the adjustment cost function C_t may also vary between periods. Starting at $x_0 = \underline{x} \in \arg \max_{x \in L} F(x, \theta)$, the agent chooses a path $(x_t)_{t=1}^\infty$ to maximize the discounted sum of her period payoffs $F(x_t, \theta_t) - C_t(x_t - x_{t-1})$. Theorem 3 validates the le Chatelier principle: under the same assumptions (ordinal complementarity of F and monotonicity of each C_t), if $\underline{\theta} \leq \theta_t \leq \bar{\theta}$ in every period t , then given any $\bar{x} \in \arg \max_{x \in L} F(x, \bar{\theta})$ such that $\underline{x} \leq \bar{x}$, the agent's choices satisfy $\underline{x} \leq x_t \leq \bar{x}$ along some optimal path $(x_t)_{t=1}^\infty$. If the parameter and cost are time-invariant ($\theta_t = \bar{\theta}$ and $C_t = C$ for all periods t), then a stronger le Chatelier principle holds (Theorem 4): under additional assumptions, the agent adjusts more at longer horizons, in the sense that $\underline{x} \leq x_t \leq x_T \leq \bar{x}$ holds at any dates $t < T$ along some optimal path $(x_t)_{t=1}^\infty$.

The le Chatelier principle remains valid if decisions are instead made by a sequence of short-lived agents (Theorem 5): under similar assumptions, if $\underline{\theta} \leq \theta_t (\leq \theta_{t+1}) \leq \bar{\theta}$ in every period t , then $\underline{x} \leq \tilde{x}_t (\leq \tilde{x}_{t+1}) \leq \bar{x}$ in every period t along some equilibrium path $(\tilde{x}_t)_{t=1}^\infty$. Thus short-lived agents adjust in the same direction as a long-lived agent would. They do so more sluggishly, however: Theorem 6 asserts that under stronger assumptions (including convexity of each cost C_t), $\underline{x} \leq \tilde{x}_t \leq x_t \leq \bar{x}$ holds along some short-lived equilibrium path $(\tilde{x}_t)_{t=1}^\infty$ and some long-lived optimal path $(x_t)_{t=1}^\infty$.

The rest of this paper is arranged as follows. In the next section, we describe the environment. We present our fundamental comparative-statics insight (Theorem 1) in section 3. In section 4, we develop a general le Chatelier principle (Theorem 2), and apply it to pricing and factor demand. In section 5, we introduce a dynamic, forward-looking adjustment model, derive two dynamic le Chatelier principles (Theorems 3 and 4), and apply them to pricing

and investment. We conclude in section 6 by deriving the le Chatelier principle for short-lived agents (Theorem 5) and comparing their behavior to that of a long-lived agent (Theorem 6). The appendix contains definitions of some standard terms, an extension to allow for uncertain adjustment costs, and all proofs omitted from the text.

2 Setting

The agent's objective is $F(x, \theta)$, where x is the choice variable and $\theta \in \Theta$ is a parameter. The choice variable x belongs to a subset L of \mathbb{R}^n .

At the initial parameter $\theta = \underline{\theta}$, an optimal choice \underline{x} was made:

$$\underline{x} \in \arg \max_{x \in L} F(x, \underline{\theta}).$$

(Note that we allow for a multiplicity of optimal actions.) This is the agent's "starting point," and we shall consider how she responds in the short and long run to a change in the parameter from $\underline{\theta}$ to $\bar{\theta}$, where $\underline{\theta} < \bar{\theta}$.

Adjustment is costly: adjusting from \underline{x} to x costs $C(x - \underline{x})$. The cost function C is a map $\Delta L \rightarrow [0, \infty]$, where $\Delta L = \{x - y : x, y \in L\}$. Note that we allow some adjustments $\epsilon \in \Delta L$ to have infinite cost $C(\epsilon) = \infty$, meaning that they are infeasible. We assume throughout that $C(0) < \infty$.

The agent adjusts her action $x \in L$ to maximize

$$G(x, \bar{\theta}) = F(x, \bar{\theta}) - C(x - \underline{x}).$$

2.1 Order assumptions

Throughout, \mathbb{R}^n (and thus L) is endowed with the usual "product" order \geq , so " $x \geq y$ " means " $x_i \geq y_i$ for every dimension i ." We write " $x > y$ " whenever $x \geq y$ and $x \neq y$. We assume that the choice set L is a *sublattice* of \mathbb{R}^n : for any $x, y \in L$, the following two vectors also belong to L :

$$\begin{aligned} x \wedge y &= (\min\{x_1, y_1\}, \dots, \min\{x_n, y_n\}) \\ \text{and } x \vee y &= (\max\{x_1, y_1\}, \dots, \max\{x_n, y_n\}). \end{aligned}$$

Examples of sublattices include $L = \mathbb{R}^n$, “boxes” $L = \{x \in \mathbb{R}^n : x_- \leq x \leq x_+\}$ for given $x_-, x_+ \in \mathbb{R}^n$, and “grids” such as $L = \mathbb{Z}^n$ (where \mathbb{Z} denotes the integers). Another example is the half-plane $L = \{(x_1, x_2) \in \mathbb{R}^2 : \alpha x_1 + \beta x_2 \geq k\}$ for given $\alpha \leq 0 \leq \beta$ and $k \in \mathbb{R}$.

The parameter θ belongs to a partially ordered set Θ . We use the symbol \geq also for this partial order. In applications, the parameter θ is often a vector of real numbers, in which case Θ is a subset of \mathbb{R}^n and \geq is the usual “product” order. More elaborate applications are possible; for example, Θ could be a set of distributions, with \geq being first-order stochastic dominance.

2.2 Monotonicity assumptions on costs

Our results feature different assumptions about the cost function C . In particular, most of our results require *monotonicity*, but our first theorem requires only *minimal monotonicity*. We now define these two properties.

The cost function C is *monotone* exactly if for any adjustment vector $\epsilon \in \Delta L$ and any dimension i ,

$$C(\epsilon_1, \dots, \epsilon_{i-1}, \epsilon'_i, \epsilon_{i+1}, \dots, \epsilon_n) \leq C(\epsilon) \quad \text{whenever } 0 \leq \epsilon'_i \leq \epsilon_i \text{ or } 0 \geq \epsilon'_i \geq \epsilon_i.$$

In other words, modifying an adjustment vector by shifting one dimension’s adjustment toward zero always reduces cost. An equivalent definition of monotonicity is that $C(\epsilon') \leq C(\epsilon)$ holds whenever ϵ' is “between 0 and ϵ ” in the sense that in each dimension i , we have either $0 \leq \epsilon'_i \leq \epsilon_i$ or $0 \geq \epsilon'_i \geq \epsilon_i$.

If C has the *additively separable* form $C(\epsilon) = \sum_{i=1}^n C_i(\epsilon_i)$, then it is monotone exactly if each of the cost functions C_i is single-dipped and minimized at zero.² Thus in particular, if the choice variable is one-dimensional ($L \subseteq \mathbb{R}$), then monotonicity requires that C be single-dipped and minimized at zero.

Example 1. For a one-dimensional choice variable ($L \subseteq \mathbb{R}$), the following cost functions are monotone, for any values of the parameters $k, a \in (0, \infty)$:
(a) Fixed cost: $C(\epsilon) = k$ for $\epsilon \neq 0$ and $C(0) = 0$. (b) Quadratic cost: $C(\epsilon) = a\epsilon^2$. (c) Quadratic cost with free disposal: $C(\epsilon) = a\epsilon^2$ if $\epsilon \geq 0$ and $C(\epsilon) = 0$

²Given $X \subseteq \mathbb{R}$, a function $\phi : X \rightarrow [0, \infty]$ is *single-dipped* exactly if there is an $x \in X$ such that ϕ is decreasing on $\{y \in X : y \leq x\}$ and increasing on $\{y \in X : y \geq x\}$.

otherwise. (d) Quadratic cost with a constraint: $C(\epsilon) = a\epsilon^2$ if $\epsilon \in \mathcal{E}$ and $C(\epsilon) = \infty$ otherwise, where the constraint set $\mathcal{E} \subseteq \mathbb{R}$ is convex and contains 0.

Example 2. The following cost functions are monotone: (a) Additively separable: $C(\epsilon) = \sum_{i=1}^n C_i(\epsilon_i)$, where each C_i is of one of the types in Example 1. (b) Euclidean: $C(\epsilon) = \sqrt{\sum_{i=1}^n \epsilon_i^2}$. (c) Cobb–Douglas: $C(\epsilon) = \prod_{i=1}^n |\epsilon_i|^{a_i}$, where $a_1, \dots, a_n \in (0, \infty)$.

Monotonicity is consistent with quite general nonconvexities, and even with failures of quasiconvexity: the cost function in Example 2(c) is monotone, but not quasiconvex (except if the choice variable is one-dimensional, i.e. $L \subseteq \mathbb{R}$).

A cost function C is called *0-monotone* if and only if

$$C(\epsilon_1, \dots, \epsilon_{i-1}, 0, \epsilon_{i+1}, \dots, \epsilon_n) \leq C(\epsilon)$$

for any adjustment vector $\epsilon \in \Delta L$ and any dimension i . This property weakens monotonicity by requiring cost to fall only when an adjustment vector is modified by *completely cancelling* the adjustment in one dimension.

An even weaker property is *minimal monotonicity*, which requires that

$$C(\epsilon \wedge 0) \leq C(\epsilon) \leq C(\epsilon \vee 0) \quad \text{for any adjustment vector } \epsilon \in \Delta L.$$

In other words, simultaneously cancelling all *upward* adjustments, by replacing all of the positive entries of an adjustment vector ϵ with zeroes, reduces cost; similarly, cancelling all *downward* adjustments reduces cost.

Although minimal monotonicity is weaker than 0-monotonicity, the two properties coincide for additively separable cost functions: both require that each C_i be minimized at zero. Thus in particular, minimal monotonicity and 0-monotonicity coincide when the choice variable is one-dimensional ($L \subseteq \mathbb{R}$): both require merely that cost be minimized at zero.

Example 3. Consider the cost function in Example 1(d), with a constraint set $\mathcal{E} \subseteq \mathbb{R}$ that contains 0 but is not convex. For instance, $\mathcal{E} = (-\infty, 0] \cup [I, \infty)$ for some $I > 0$, as in the recent literature on the investment behavior of entrepreneurs in developing countries (see section 5.4). Or $\mathcal{E} = \mathbb{Z}$ due to an integer constraint. Such a cost function is 0-monotone, but not monotone.

Example 4. For $L \subseteq \mathbb{R}^2$, the (nonseparable) cost function $C(\epsilon_1, \epsilon_2) = |\epsilon_1 - \epsilon_2|$ is minimally monotone, but not 0-monotone (hence not monotone).³ This captures situations in which it is costly to adjust the gap between x_1 and x_2 .

Monotonicity, 0-monotonicity and minimal monotonicity are all preserved by strictly increasing transformations: if C has one of these properties, then so does $\phi \circ C$ for any strictly increasing map $\phi : [0, \infty] \rightarrow [0, \infty]$. In other words, these properties are *ordinal*. That makes them easier to check in practice.⁴

2.3 Complementarity assumptions on the objective

We assume throughout that the objective function F satisfies the standard *ordinal complementarity conditions* of *quasi-supermodularity* and *single-crossing differences* (see Milgrom & Shannon, 1994), defined as follows.

The objective $F(x, \theta)$ has *single-crossing differences* in (x, θ) exactly if $F(y, \theta') - F(x, \theta') \geq (>) 0$ implies $F(y, \theta'') - F(x, \theta'') \geq (>) 0$ whenever $x \leq y$ and $\theta' \leq \theta''$. Economically, this means that a higher parameter implies a greater liking for higher actions: whenever a higher action is (strictly) preferred to a lower one, this remains true if the parameter increases. A sufficient condition is *increasing differences*, which requires that $F(y, \theta) - F(x, \theta)$ be increasing in θ whenever $x \leq y$. Related concepts, such as *log increasing differences*, are defined in the appendix.

A function $\phi : L \rightarrow \mathbb{R}$ is called *quasi-supermodular* if $\phi(x) - \phi(x \wedge y) \geq (>) 0$ implies $\phi(x \vee y) - \phi(y) \geq (>) 0$. A sufficient condition is *supermodularity*, which requires that $\phi(x) - \phi(x \wedge y) \leq \phi(x \vee y) - \phi(y)$ for any $x, y \in L$. If $L \subseteq \mathbb{R}$, then every function $\phi : L \rightarrow \mathbb{R}$ is automatically supermodular. In case L is open and convex and ϕ is continuously differentiable, supermodularity demands precisely that $\partial\phi/\partial x_i$ be increasing in x_j , for all dimensions $i \neq j$. See the appendix for definitions of related concepts, such as *submodularity*.

³0-monotonicity fails since, for example, $C(0, 1) = 1 > 0 = C(1, 1)$. For minimal monotonicity, if $\epsilon_1 \geq 0 \leq \epsilon_2$ then $C(\epsilon \wedge 0) = C(0, 0) = 0 \leq C(\epsilon) = C(\epsilon \vee 0)$, and similarly if $\epsilon_1 < 0 > \epsilon_2$; if $\epsilon_1 < 0 \leq \epsilon_2$ then $C(\epsilon \wedge 0) = C(\epsilon_1, 0) = |\epsilon_1| > C(0, 0) \leq |\epsilon_2| \leq C(0, \epsilon_2) = C(\epsilon \vee 0)$, and similarly if $\epsilon_1 \geq 0 > \epsilon_2$.

⁴For instance, the cost function in Example 2(b) is monotone because it is a strictly increasing transformation of the additively separable cost function $C^\dagger(\epsilon) = \sum_{i=1}^n \epsilon_i^2$, which is monotone since $\epsilon_i \mapsto \epsilon_i^2$ is single-dipped and minimized at zero.

We say that $F(x, \theta)$ is *(quasi-)supermodular in x* if for each parameter $\theta \in \Theta$, the function $F(\cdot, \theta) : L \rightarrow \mathbb{R}$ is (quasi-)supermodular. This captures complementarity between the different dimensions of the action.

3 Comparative statics

Recall that the agent chooses $x \in L$ to maximize $G(x, \bar{\theta}) = F(x, \bar{\theta}) - C(x - \underline{x})$. Our fundamental comparative-statics result is the following.

Theorem 1. *Suppose that the objective $F(x, \theta)$ is quasi-supermodular in x and has single-crossing differences in (x, θ) , and that the adjustment cost C is minimally monotone. If $\bar{\theta} > \underline{\theta}$, then $\hat{x} \geq \underline{x}$ for some $\hat{x} \in \arg \max_{x \in L} G(x, \bar{\theta})$, provided the argmax is nonempty.*

In words, an increased parameter leads to a higher action (modulo tie-breaking). This parallels the basic comparative-statics result for costless adjustment (see Milgrom & Shannon, 1994, Theorem 4), one version of which states that under the same ordinal complementarity conditions on the objective F , we have $\bar{x} \geq \underline{x}$ for some $\bar{x} \in \arg \max_{x \in L} F(x, \bar{\theta})$, provided the argmax is nonempty. Theorem 1 shows that this basic result is strikingly robust to adjustment costs: the objective F need not satisfy any additional property, and the cost C need only be minimally monotone.

Theorem 1 does *not* follow from applying the basic comparative-statics result to the objective function $G(x, \bar{\theta})$, because its assumptions do not guarantee that $G(\cdot, \bar{\theta})$ is quasi-supermodular.⁵ A different argument is required.

Proof. Let $x' \in \arg \max_{x \in L} G(x, \bar{\theta})$. We claim that $\hat{x} = \underline{x} \vee x'$ also maximizes $G(\cdot, \bar{\theta})$; obviously $\hat{x} \geq \underline{x}$. We have $F(\underline{x}, \underline{\theta}) \geq F(\underline{x} \wedge x', \underline{\theta})$ by definition of \underline{x} . Thus $F(\underline{x} \vee x', \underline{\theta}) \geq F(x', \underline{\theta})$ by quasi-supermodularity, whence $F(\underline{x} \vee x', \bar{\theta}) \geq F(x', \bar{\theta})$ by single-crossing differences. Furthermore, by minimal monotonicity, $C(\underline{x} \vee x' - \underline{x}) = C((x' - \underline{x}) \vee 0) \leq C(x' - \underline{x})$. Thus

$$G(\hat{x}, \bar{\theta}) = F(\underline{x} \vee x', \bar{\theta}) - C(\underline{x} \vee x' - \underline{x}) \geq F(x', \bar{\theta}) - C(x' - \underline{x}) = G(x', \bar{\theta}).$$

⁵Its second term $x \mapsto -C(x - \underline{x})$ need not be quasi-supermodular, and in any case, the sum of two quasi-supermodular functions is not quasi-supermodular in general.

Since x' maximizes $G(\cdot, \bar{\theta})$ on L , it follows that \hat{x} does, too.

QED

The minimal-monotonicity assumption on the cost C is essential for Theorem 1: without this assumption, it may be that $\hat{x} \not\preceq \underline{x}$.⁶

In applications, it is often useful that Theorem 1 requires F to satisfy only the ordinal complementarity conditions, rather than the stronger *cardinal complementarity conditions* of supermodularity and increasing differences. In monopoly pricing, for example, the objective F has single-crossing differences, but not increasing differences—see section 4.2 below.

Theorem 1 has a counterpart for parameter *decreases*: under the same assumptions, if $\bar{\theta} < \underline{\theta}$, then $\hat{x} \leq \underline{x}$ for some $\hat{x} \in \arg \max_{x \in L} G(x, \bar{\theta})$, provided the argmax is nonempty. The proof is exactly analogous.⁷ All of the results in this paper have such counterparts for parameter decreases; we will not discuss them explicitly.

Theorem 1 also extends straightforwardly the case in which adjustment costs are uncertain, even if the agent is risk-averse. Several subsequent results also generalize in this way. This is shown in appendix B.

Remark 1. Theorem 1 is phrased differently than the usual statement of the basic result (see Milgrom & Shannon, 1994, Theorem 4), which asserts that when $F(x, \theta)$ is quasi-supermodular in x and has single-crossing differences in (x, θ) , if $\bar{\theta} > \underline{\theta}$ then

$$\begin{aligned} x'' \in \arg \max_{x \in L} F(x, \underline{\theta}) \quad \text{and} \quad x' \in \arg \max_{x \in L} F(x, \bar{\theta}) \\ \implies x'' \wedge x' \in \arg \max_{x \in L} F(x, \underline{\theta}) \quad \text{and} \quad x'' \vee x' \in \arg \max_{x \in L} F(x, \bar{\theta}). \end{aligned}$$

A version of Theorem 1 with this form also holds: under the same hypotheses, if $\bar{\theta} > \underline{\theta}$ then

$$\begin{aligned} \underline{x} \in \arg \max_{x \in L} F(x, \underline{\theta}) \quad \text{and} \quad x' \in \arg \max_{x \in L} G(x, \bar{\theta}) \\ \implies \underline{x} \wedge x' \in \arg \max_{x \in L} F(x, \underline{\theta}) \quad \text{and} \quad \underline{x} \vee x' \in \arg \max_{x \in L} G(x, \bar{\theta}). \end{aligned}$$

⁶For example, if $C(\epsilon) = \infty$ for all $\epsilon \neq \hat{\epsilon}$ and $C(\hat{\epsilon}) = 0$, where $\hat{\epsilon} \not\preceq 0$, then $\hat{x} = \underline{x} + \hat{\epsilon} \not\preceq \underline{x}$.

⁷The proof of Theorem 1 uses only one-half of the minimal monotonicity assumption: that $C(\epsilon \vee 0) \leq C(\epsilon)$ for every $\epsilon \in \Delta L$. The proof of its parameter-decrease counterpart uses (only) the other half, namely that $C(\epsilon \wedge 0) \leq C(\epsilon)$ for every $\epsilon \in \Delta L$.

The latter property (concerning $\underline{x} \vee x'$) is exactly what is shown in the proof of Theorem 1. For the former property (concerning $\underline{x} \wedge x'$), suppose it were to fail; then $F(\underline{x}, \underline{\theta}) > F(\underline{x} \wedge x', \underline{\theta})$, so that replicating the steps in the proof of Theorem 1 delivers $G(\underline{x} \vee x', \bar{\theta}) > G(x', \bar{\theta})$, a contradiction with the fact that x' maximizes $G(\cdot, \bar{\theta})$ on L . \triangle

3.1 “ \forall ” comparative statics

We now provide a “ \forall ” counterpart to Theorem 1, giving two conditions under either of which $\hat{x} \geq \underline{x}$ holds for *every* optimal choice \hat{x} . The first of these conditions is *strict single-crossing differences* of the objective $F(x, \theta)$ in (x, θ) , which requires that $F(y, \theta') - F(x, \theta') \geq 0$ implies $F(y, \theta'') - F(x, \theta'') > 0$ whenever $x < y$ and $\theta' < \theta''$. The second is *strict minimal monotonicity* of the cost C , which demands that for any adjustment vector $\epsilon \in \Delta L$,

$$C(\epsilon \wedge 0) < C(\epsilon) \quad \text{unless } \epsilon \leq 0, \quad \text{and} \quad C(\epsilon \vee 0) < C(\epsilon) \quad \text{unless } \epsilon \geq 0.^8$$

In other words, simultaneously cancelling all upward adjustments, by replacing all of the strictly positive entries of an adjustment vector ϵ with zeroes, *strictly* reduces cost; and likewise for downward adjustments.

Proposition 1. *Suppose that the objective $F(x, \theta)$ is quasi-supermodular in x , and that either*

- (a) *the objective $F(x, \theta)$ has strict single-crossing differences in (x, θ) and the cost C is minimally monotone, or*
- (b) *the objective $F(x, \theta)$ has single-crossing differences in (x, θ) and the cost C is strictly minimally monotone.*

*If $\bar{\theta} > \underline{\theta}$, then $\hat{x} \geq \underline{x}$ for any $\hat{x} \in \arg \max_{x \in L} G(x, \bar{\theta})$.*⁹

Proposition 1 is the costly-adjustment analog of the standard “ \forall ” comparative-statics result (see Milgrom & Shannon, 1994, Theorem 4'), which

⁸Equivalently: $C(\epsilon \wedge 0) < C(\epsilon)$ unless $\epsilon \wedge 0 = \epsilon$, and $C(\epsilon \vee 0) < C(\epsilon)$ unless $\epsilon \vee 0 = 0$.

⁹A variant of Proposition 1 can be obtained by mixing the “strictness” properties (a) and (b): if $x = (y, z)$, where $F(y, z, \theta)$ has strict single-crossing differences in (y, θ) for any fixed z , and $C(y - \underline{y}, \cdot)$ is strictly minimally monotone for any fixed y , then the conclusion goes through, with essentially the same proof.

states that given any $\underline{x} \in \arg \max_{x \in L} F(x, \underline{\theta})$, if $F(x, \theta)$ is quasi-supermodular in x and has strict single-crossing differences in (x, θ) , then $\bar{x} \geq \underline{x}$ for *any* $\bar{x} \in \arg \max_{x \in L} F(x, \bar{\theta})$. Part (a) directly extends this result to the costly-adjustment case. Part (b) shows that the “strictness” in the hypotheses required to obtain a “ \forall ” comparative-statics conclusion can come from the cost C rather than the objective F : in particular, strict minimal monotonicity ensures that even if some action $x \not\geq \underline{x}$ maximizes $F(\cdot, \bar{\theta})$, it will not be chosen due to its cost.

4 The le Chatelier principle

The le Chatelier principle asserts that long-run elasticities exceed short-run elasticities. In this section, we show that the le Chatelier principle is far more general than previously claimed: it arises whenever adjustment is costly, provided merely that the cost function is monotone. The classic formalization, which assumes that only some dimensions of the action are adjustable in the short run, is the special case in which each dimension has an adjustment cost that is either prohibitively high or equal to zero.

We consider the agent’s short- and long-run responses to a shift of the parameter from $\underline{\theta}$ to $\bar{\theta}$. Her short-run response \hat{x} takes adjustment costs into account, so it maximizes $G(\cdot, \bar{\theta})$. In the long run, the agent adjusts to a new frictionless optimum $\bar{x} \in \arg \max_{x \in L} F(x, \bar{\theta})$.

Recall from section 2.2 the definition of a *monotone* cost function C .

Theorem 2 (le Chatelier principle). *Suppose that the objective $F(x, \theta)$ is quasi-supermodular in x and has single-crossing differences in (x, θ) , and that the adjustment cost C is monotone. Fix $\bar{\theta} > \underline{\theta}$, and let $\bar{x} \in \arg \max_{x \in L} F(x, \bar{\theta})$ satisfy $\bar{x} \geq \underline{x}$.¹⁰ Then*

- $\bar{x} \geq \hat{x} \geq \underline{x}$ for some $\hat{x} \in \arg \max_{x \in L} G(x, \bar{\theta})$, provided the argmax is nonempty, and
- if \bar{x} is the largest element of $\arg \max_{x \in L} F(x, \bar{\theta})$, then $\bar{x} \geq \hat{x}$ for any $\hat{x} \in \arg \max_{x \in L} G(x, \bar{\theta})$.

¹⁰Such an \bar{x} must exist, provided the argmax is nonempty. This follows from the basic comparative-statics result (see Milgrom & Shannon, 1994, Theorem 4).

Theorem 2 nests the le Chatelier principle of Milgrom and Roberts (1996), in which it is assumed that only some dimensions x_i of the choice variable can be adjusted in the short run, and that such adjustments are costless. This is the special case of our model in which $C(\epsilon) = \sum_{i=1}^n C_i(\epsilon_i)$, where some dimensions i have $C_i \equiv 0$, and the other dimensions i have $C_i(\epsilon_i) = \infty$ for every $\epsilon_i \neq 0$.

Like Theorem 1, Theorem 2 requires F only to satisfy ordinal complementarity properties, not cardinal ones. This greatly extends its applicability, allowing it to be used to study pricing, for example (see section 4.2 below).

Proof. For the first part, assume that $\arg \max_{x \in L} G(x, \bar{\theta})$ is nonempty. By Theorem 1, we may choose an $x' \in \arg \max_{x \in L} G(x, \bar{\theta})$ such that $x' \geq \underline{x}$. We claim that $\hat{x} = \bar{x} \wedge x'$ also maximizes $G(\cdot, \bar{\theta})$; this suffices since $\bar{x} \geq \bar{x} \wedge x' \geq \underline{x}$. We have $F(\bar{x} \vee x', \bar{\theta}) \leq F(\bar{x}, \bar{\theta})$ by definition of \bar{x} , which by quasi-supermodularity implies that $F(x', \bar{\theta}) \leq F(\bar{x} \wedge x', \bar{\theta})$. Since C is monotone and $x' \geq \bar{x} \wedge x' \geq \underline{x}$, we have $C(x' - \underline{x}) \geq C(\bar{x} \wedge x' - \underline{x})$. Thus

$$G(x', \bar{\theta}) = F(x', \bar{\theta}) - C(x' - \underline{x}) \leq F(\bar{x} \wedge x', \bar{\theta}) - C(\bar{x} \wedge x' - \underline{x}) = G(\hat{x}, \bar{\theta}),$$

which since x' maximizes $G(\cdot, \bar{\theta})$ on L implies that \hat{x} does, too.

For the second part, let \bar{x} be the largest element of $\arg \max_{x \in L} F(x, \bar{\theta})$, and let $\hat{x} \in \arg \max_{x \in L} G(x, \bar{\theta})$; we will show that $\bar{x} \geq \hat{x}$. The optimality of \hat{x} implies that $G(\hat{x}, \bar{\theta}) \geq G(\bar{x} \wedge \hat{x}, \bar{\theta})$. It furthermore holds that $C(\hat{x} - \underline{x}) \geq C(\bar{x} \wedge \hat{x} - \underline{x})$, by the monotonicity of C and the fact that in each dimension i , either $\bar{x}_i \leq \hat{x}_i$ so $0 \leq (\bar{x} \wedge \hat{x} - \underline{x})_i \leq (\hat{x} - \underline{x})_i$, or $\bar{x}_i > \hat{x}_i$ in which case $(\bar{x} \wedge \hat{x} - \underline{x})_i = (\hat{x} - \underline{x})_i$. Hence $F(\hat{x}, \bar{\theta}) \geq F(\bar{x} \wedge \hat{x}, \bar{\theta})$, which implies $F(\bar{x} \vee \hat{x}, \bar{\theta}) \geq F(\bar{x}, \bar{\theta})$ by quasi-supermodularity. Since \bar{x} is the largest maximizer of $F(\cdot, \bar{\theta})$, it follows that $\bar{x} \geq \bar{x} \vee \hat{x}$, which is to say that $\bar{x} \geq \hat{x}$. **QED**

The monotonicity assumption in Theorem 2 is essential: if the cost C were merely minimally monotone, then $\bar{x} \geq \hat{x}$ would not necessarily hold.¹¹

Theorem 2 remains true if adjustment is costly also in the long run: that is, if in addition to the short-run cost $C_1(x_1 - \underline{x})$ of moving from the initial

¹¹For example, if $L = \mathbb{R}$, $F(x, \bar{\theta}) = -x^2$, $F(x, \bar{\theta}) = -(x-2)^2$, and $C(\epsilon) = \infty$ if $0 < \epsilon < 3$ and $C(\epsilon) = 0$ otherwise, then $G(\cdot, \bar{\theta})$ is uniquely maximized by $\hat{x} = 3$, and $\bar{x} = 2 < \hat{x}$.

choice \underline{x} to her short-run choice x_1 , the agent incurs a further cost $C_2(x_2 - x_1)$ of moving from her short-run choice x_1 to her long-run choice x_2 .

Proposition 2. *Suppose that the objective $F(x, \theta)$ is quasi-supermodular in x and has single-crossing differences in (x, θ) , and that the adjustment costs C_1 and C_2 are monotone. If $\bar{\theta} > \underline{\theta}$, then $x_2 \geq x_1 \geq \underline{x}$ for some $x_1 \in \arg \max_{x \in L} [F(x, \bar{\theta}) - C_1(x - \underline{x})]$ and $x_2 \in \arg \max_{x \in L} [F(x, \bar{\theta}) - C_2(x - x_1)]$, provided the argmaxes are nonempty.¹²*

Proposition 2 is a special case of a later result, Theorem 5 (section 6 below).

4.1 Application to factor demand

Consider a stylized model of production, following Milgrom and Roberts (1996). A firm uses capital k and labor ℓ to produce output $f(k, \ell)$. Profit at real factor prices (r, w) is $F(k, \ell, -w) = f(k, \ell) - rk - w\ell$. The adjustment cost C is monotone, but otherwise unrestricted.

If the production function f is supermodular, meaning that capital and labor are complements, then profit $F(k, \ell, -w)$ is supermodular in $x = (k, \ell)$. By inspection, the profit function $F(k, \ell, -w)$ has increasing differences in $(x, \theta) = ((k, \ell), -w)$. So by Theorem 2, any drop in the wage w precipitates a short-run increase of both k and ℓ , and a further increase in the long run.

If f is instead submodular, meaning that capital and labor are substitutes in production, then we may apply Theorem 2 to the choice variable $(x_1, x_2) = (-k, \ell)$, since profit $F^\dagger(x_1, x_2, -w) = f(-x_1, x_2) + rx_1 + (-w)x_2$ is then supermodular in $x = (x_1, x_2)$ and has increasing differences in $(x, -w)$.¹³ The conclusion is that ℓ still increases in the short run and further increases in the long run, whereas k now *decreases*.

Milgrom and Roberts (1996) were the first to use the theory of monotone comparative statics to obtain such a result. They assumed that labor adjustments are costless and that capital cannot be adjusted at all in the short run: in other words, $C(\epsilon_k, \epsilon_\ell) = C_k(\epsilon_k) + C_\ell(\epsilon_\ell)$, where $C_\ell \equiv 0$ and $C_k(\epsilon_k) = \infty$ for every $\epsilon_k \neq 0$. Our analysis reveals that much weaker assumptions suffice. It

¹²In fact, x_1 and x_2 may be chosen so that $\bar{x} \geq x_2 \geq x_1 \geq \underline{x}$ holds for any $\bar{x} \in \arg \max_{x \in L} F(x, \bar{\theta})$ that satisfies $\bar{x} \geq \underline{x}$.

¹³This trick is due to Milgrom and Roberts (1996).

turns out not to matter whether labor is cheap to adjust relative to capital. What matters is, rather, that short-run adjustments are costly.

4.2 Application to pricing

The central plank of new Keynesian macroeconomic models is price stickiness, and the oldest and most important microfoundation for this property is (nonconvex) adjustment costs (e.g. Mankiw, 1985; Caplin & Spulber, 1987; Golosov & Lucas, 2007; Midrigan, 2011). These may be real costs of updating what prices are displayed: empirically, such “menu costs” can be nonnegligible (see e.g. Levy, Bergen, Dutta, & Venable, 1997). Or they may arise from consumers reacting adversely to price hikes by temporarily reducing demand (as in Antić & Salant, in progress).

To study pricing, we consider the simplest model, following Milgrom and Roberts (1990): a monopolist with constant marginal cost $c \geq 0$ faces a decreasing demand curve $D(\cdot, \eta)$ parametrized by η , thus earning a profit of $F(p, (c, -\eta)) = (p - c)D(p, \eta)$ if she prices at $p \in \mathbb{R}_+$. We assume that demand $D(p, \eta)$ is always strictly positive, and that η is an elasticity shifter: when it increases, so does the absolute elasticity of demand at every price p . Then profit $F(p, \theta) = F(p, (c, -\eta))$ has increasing differences in (p, c) and has log increasing differences in $(p, -\eta)$, so it has single-crossing differences in $(p, \theta) = (p, (c, -\eta))$. Furthermore, profit $F(p, \theta)$ is automatically quasi-supermodular in p since this choice variable is one-dimensional ($L \subseteq \mathbb{R}$).

Adjusting the price by ϵ incurs a cost of $C(\epsilon) \geq 0$. We assume nothing about C except that it is minimized at zero. In many macroeconomic models, it is a pure fixed cost: $C(\epsilon) = k > 0$ for every $\epsilon \neq 0$. When adjustment costs arise from price-hike-averse consumers, we have $C(\epsilon) = 0$ for $\epsilon \leq 0$ and $C(\epsilon) > 0$ for $\epsilon > 0$. If consumers are inattentive to small price changes, then $C(\epsilon) = 0$ if $\epsilon \in [\underline{\epsilon}, \bar{\epsilon}]$ and $C(\epsilon) > 0$ otherwise, where $\underline{\epsilon} < 0 < \bar{\epsilon}$.

By Theorem 1, the familiar comparative-statics properties of the monopoly problem are robust to the introduction of adjustment costs: it remains true that the monopolist raises her price whenever her marginal cost c rises and whenever demand becomes less elastic (i.e., η falls). No assumptions on the adjustment cost C are required except that it be minimized at zero.

Under the mild additional assumption that C is single-dipped, Theorem 2 yields a dynamic prediction: in response to a shock that increases her marginal cost or decreases the elasticity of demand, the monopolist initially raises her price, and then increases it further over the longer run. Thus one-off permanent cost and demand-elasticity shocks lead, quite generally, to price increases in both the short and long run.

A key reason why we can draw such general conclusions about pricing is that Theorems 1 and 2 require F to satisfy only ordinal (not cardinal) complementarity conditions. Specifically, we used the fact that the monopolist's profit undergoes a “single-crossing differences” shift when demand becomes less elastic (i.e., when η falls). A result which assumed the cardinal property of *increasing* differences would have been inapplicable, since elasticity shifts do not generally cause profit to shift in an “increasing differences” fashion.¹⁴

5 Dynamic adjustment

The le Chatelier principle takes a classical, “reduced-form” approach to dynamics, following Samuelson and Milgrom–Roberts. In this section and the next, we consider a fully-fledged dynamic model of adjustment. We show that the le Chatelier principle remains valid: in the short run, the agent's choices exceed the initial choice \underline{x} and do not overshoot the new frictionless optimum \bar{x} . We furthermore show that under additional assumptions, the path of adjustment is monotone, so that the agent adjusts more over longer horizons.

In this section, we assume that the agent is long-lived and forward-looking. The alternative case in which each period t 's choice x_t is made by a short-lived agent (or equivalently, by a myopic long-lived agent) is studied in section 6.

5.1 Setting

The agent faces an infinite-horizon decision problem in discrete time. In each period $t \in \mathbb{N} = \{1, 2, 3, \dots\}$, she takes an action $x_t \in L$, and earns a payoff of

¹⁴This applied advantage of requiring only ordinal complementarity was emphasized by Milgrom and Roberts (1990) and Milgrom and Shannon (1994) in the context of models with costless adjustment.

$F(x_t, \theta_t)$. Adjusting from x_{t-1} to x_t in period t costs $C_t(x_t - x_{t-1})$. The agent's initial choice $x_0 = \underline{x} \in \arg \max_{x \in L} F(x, \underline{\theta})$ is given.

The parameter sequence $(\theta_t)_{t=1}^\infty$ is given, as is the sequence $(C_t)_{t=1}^\infty$ of adjustment cost functions. The simplest example is a one-off parameter shift ($\theta_t = \bar{\theta}$ for all $t \in \mathbb{N}$) with a time-invariant cost ($C_t = C$ for all $t \in \mathbb{N}$).

The agent is forward-looking, and discounts future payoffs by a factor of $\delta \in (0, 1)$. Given her period-0 choice $x_0 \in L$, the agent's payoff from a sequence $(x_t)_{t=1}^\infty$ in L is

$$\mathcal{G}((x_t)_{t=1}^\infty, x_0) = \mathcal{F}((x_t)_{t=1}^\infty) - \mathcal{C}(x_0, (x_t)_{t=1}^\infty), \quad \text{where}$$

$$\mathcal{F}((x_t)_{t=1}^\infty) = \sum_{t=1}^\infty \delta^{t-1} F(x_t, \theta_t) \quad \text{and} \quad \mathcal{C}(x_0, (x_t)_{t=1}^\infty) = \sum_{t=1}^\infty \delta^{t-1} C_t(x_t - x_{t-1}).$$

5.2 Dynamic le Chatelier principles

The following result shows that our le Chatelier principle (Theorem 2) remains valid when the agent can adjust over time and is forward-looking: for any new frictionless optimum $\bar{x} \in \arg \max_{x \in L} F(x, \bar{\theta})$, the agent's "short-run" actions x_t satisfy $\underline{x} \leq x_t \leq \bar{x}$ along some optimal path $(x_t)_{t=1}^\infty$.

Theorem 3 (dynamic le Chatelier). *Suppose that the objective $F(x, \theta)$ is quasi-supermodular in x and has single-crossing differences in (x, θ) , and that each adjustment cost C_t is monotone. Fix $\bar{\theta} > \underline{\theta}$, and let $\bar{x} \in \arg \max_{x \in L} F(x, \bar{\theta})$ satisfy $\bar{x} \geq \underline{x}$.¹⁵ If $\underline{\theta} \leq \theta_t \leq \bar{\theta}$ for every $t \in \mathbb{N}$, then provided the long-lived agent's problem admits a solution, there is a solution $(x_t)_{t=1}^\infty$ that satisfies $\underline{x} \leq x_t \leq \bar{x}$ for every period $t \in \mathbb{N}$.*

The proof (appendix E) is a direct extension of the arguments used to prove Theorems 1 and 2. The straightforwardness of this extension is perhaps surprising, since the dynamic adjustment problem is superficially quite different from the one-shot problem: the agent chooses a *sequence* of actions, and her objective $\mathcal{F}(\cdot)$ need not be quasi-supermodular (since the sum of quasi-supermodular functions is not quasi-supermodular in general).

The next result shows that under stronger assumptions, a stronger dynamic le Chatelier principle holds: $\underline{x} \leq x_t \leq x_T \leq \bar{x}$ for any periods $t < T$, which

¹⁵Such an \bar{x} must exist, provided the argmax is nonempty (refer to footnote 10).

is to say that the agent adjusts more at longer horizons. Let us use “BCS” as shorthand for “bounded on compact sets.”

Theorem 4 (strong dynamic le Chatelier). *Suppose that the objective $F(x, \theta)$ is supermodular and BCS in x and has single-crossing differences in (x, θ) , and that $C_t = C$ for every period t , where the adjustment cost C is monotone and additively separable. Fix $\bar{\theta} > \underline{\theta}$, and let $\bar{x} \in \arg \max_{x \in L} F(x, \bar{\theta})$ satisfy $\bar{x} \geq \underline{x}$.¹⁶ Let the parameter shift once and for all: $\theta_t = \bar{\theta}$ for every $t \in \mathbb{N}$. Then provided the long-lived agent’s problem admits a solution, there is a solution $(x_t)_{t=1}^\infty$ that satisfies $\underline{x} \leq x_t \leq x_{t+1} \leq \bar{x}$ for every period $t \in \mathbb{N}$.*

The assumptions of Theorem 4 strengthen those of Theorem 3 in two directions: (1) the parameter θ_t and cost function C_t must not vary over time, and (2) the payoff F and cost $C_t = C$ must satisfy additional properties, namely, supermodularity and BCS of $F(\cdot, \theta)$ and additive separability of C . The BCS requirement is mild (continuity is a sufficient condition). The other two requirements, supermodularity and additive separability, have more substance. However, they are both automatically satisfied when the choice variable is one-dimensional ($L \subseteq \mathbb{R}$), as in our applications to pricing and capital investment (sections 5.3 and 5.4 below).

The proof (appendix F) runs as follows. Any sequence $(x_t)_{t=1}^\infty$ can be made increasing by replacing its t^{th} entry x_t with the cumulative maximum $x_1 \vee x_2 \vee \cdots \vee x_{t-1} \vee x_t$, for each $t \in \mathbb{N}$. It suffices to show that “monotonization” preserves optimality, since then the optimal sequence delivered by Theorem 3 may be monotonized to yield an optimal sequence with all of the desired properties. To prove that monotonization preserves optimality, it suffices (by BCS and a limit argument) to show that an optimal sequence $(x_t)_{t=1}^\infty$ which satisfies $x_1 \leq x_2 \leq \cdots \leq x_{k-1} \leq x_k$ remains optimal if its t^{th} entry x_t is replaced by $x_{t-1} \vee x_t$ for each $t \geq k+1$. We prove this using the supermodularity of $F(\cdot, \bar{\theta})$ and the monotonicity and additive separability of C .

Remark 2. There is familiar way of obtaining comparative statics in dynamic problems, via the Bellman equation (see Hopenhayn & Prescott, 1992). This approach can be used to obtain an increasing optimal path $(x_t)_{t=1}^\infty$, as in The-

¹⁶Such an \bar{x} must exist, provided the argmax is nonempty (refer to footnote 10).

orem 4, but only under stronger assumptions. The Bellman equation is

$$V(x) = \max_{x' \in L} [F(x', \bar{\theta}) - C(x' - x) + \delta V(x')] \quad \text{for every } x \in L.$$

Suppose we find conditions which guarantee that

$$\Phi(x) = \arg \max_{x' \in L} [F(x', \bar{\theta}) - C(x' - x) + \delta V(x')]$$

is a nonempty compact sublattice for each $x \in L$, and that the correspondence Φ is increasing in the strong set order. Then $\Phi(x)$ has a greatest element $\phi(x)$ for each $x \in L$, and ϕ is an increasing function. Define an optimal sequence $(x_t)_{t=1}^\infty$ by $x_t = \phi(x_{t-1})$ for each $t \in \mathbb{N}$. We have $x_1 = \phi(x_0) \geq x_0$ by Theorem 3, and by repeatedly applying ϕ on both sides of this inequality, we obtain $x_2 = \phi(x_1) \geq x_1$, then $x_3 = \phi(x_2) \geq x_2$, and so on; thus $(x_t)_{t=1}^\infty$ is increasing.

The usual sufficient conditions for the above argument are that $G(x', x) = F(x', \bar{\theta}) - C(x' - x)$ is supermodular in x' and has increasing differences in (x', x) . These cardinal complementarity conditions, together with ancillary assumptions, ensure that the value function V is supermodular (see Hopenhayn & Prescott, 1992), allowing us to apply the basic comparative-statics result (see Milgrom & Shannon, 1994, Theorem 4) to the objective $G(x', x) + \delta V(x')$ to conclude that Φ is increasing in the strong set order. The cardinal complementarity conditions effectively require that $-C(x' - x)$ have increasing differences (x', x) . This is a very restrictive assumption; in the additively separable case $C(\epsilon) = \sum_{i=1}^n C_i(\epsilon_i)$, it demands that each C_i be convex. Theorem 4 avoids this assumption, requiring merely that the cost C be monotone. \triangle

5.3 Application to pricing, continued

An active literature in macroeconomics (e.g. Golosov & Lucas, 2007; Midrigan, 2011) examines the price stickiness central to the new Keynesian paradigm by studying forward-looking dynamic models of pricing subject to adjustment costs (usually called “menu costs” in this context—see section 4.2). The basic mechanism is that nonconvexities in adjustment costs give rise to price stickiness.

Theorem 4 delivers comparative statics for such pricing models, without

any of the parametric assumptions that are typically placed on adjustment costs.¹⁷ Consider again the monopoly pricing problem described in section 4.2. Assume that the adjustment cost C is monotone. It is automatically true that the cost C is additively separable and that profit $F(\cdot, \eta)$ is supermodular, because the choice variable $p \in \mathbb{R}_+$ is one-dimensional. Thus by Theorem 4, supply shocks cause inflation at every horizon: a one-off permanent increase of marginal cost c leads prices to increase monotonically over time. The same is true of demand shocks that make the demand curve less elastic (lower η).

Theorem 3 furthermore provides that the path of prices remains always above the original frictionless monopoly price, and never overshoots the new frictionless monopoly price. This conclusion is more general, holding even when marginal cost c_t , the demand elasticity parameter η_t , and the adjustment cost function $C_t(\cdot)$ vary over time.

Although we phrased these findings in terms of a monopolist's pricing problem, they apply equally to the typical new Keynesian setting of monopolistic competition between many firms selling differentiated goods (see e.g. Galí, 2015). In that case, the demand curve in our analysis above is to be understood as *residual* demand, taking into account the other firms' pricing.

5.4 Application to capital investment

In the neoclassical theory of investment (originating with Jorgenson, 1963), a firm adjusts its capital stock over time subject to adjustment costs. In the simplest such model, the profit of a firm with capital stock $k_t \in \mathbb{R}_+$ is $F(k_t, (p, \eta, -r)) = pf(k_t, \eta) - rk_t$, where (p, r) are the prices of output and capital and $f(\cdot, \eta)$ is an increasing production function. Capital is subject to an adjustment cost: investing $i_t = k_t - k_{t-1}$ costs $C(i_t) \geq 0$, where $C(0) = 0$.

We assume that f has increasing differences, so that the parameter η shifts the marginal product of capital. Then $F(k, \theta)$ has increasing differences (and hence single-crossing differences) in (k, θ) , where $\theta = (p, \eta, -r)$. Profit $F(k, \theta)$ is automatically supermodular in k , since $k \in \mathbb{R}_+$ is one-dimensional. Our discussion below may be extended to richer variants of this model featuring, for example, depreciation and time-varying prices.

¹⁷Common functional forms include quadratic (Rotemberg, 1982) and pure fixed cost (many papers, e.g. Caplin & Spulber, 1987; Golosov & Lucas, 2007; Midrigan, 2011).

The early literature assumed a convex adjustment cost $C(\cdot)$, which yields gradual capital accumulation and an equivalence of the neoclassical theory with Tobin’s (1969) “ q ” theory of investment (see Hayashi, 1982). Later work focused on the “lumpy” investment behavior that arises when adjustment costs are nonconvex. “Lumpiness” is empirically well-documented (see Cooper & Haltiwanger, 2006), and has implications for, among other things, business cycles (e.g. Thomas, 2002; Bachmann, Caballero, & Engel, 2013; Winberry, 2021) and the effects of microfinance programs on entrepreneurship in developing countries (e.g. Field, Pande, Papp, & Rigol, 2013; Bari, Malik, Meki, & Quinn, 2021).

Our comparative-statics theory handles both the convex case and rich forms of nonconvexity. Our le Chatelier principles (the “classical,” reduced-form Theorem 2 and the dynamic, forward-looking Theorems 3 and 4) are applicable provided merely that adjustment costs are single-dipped. Investment then increases at every horizon, and by more at longer horizons, whenever the marginal profitability of capital increases, whether due to a drop in its price r , a rise in the price p of output, or an increase of the marginal product of capital (an increase of η).

The aforementioned papers on microfinance consider models in which adjustment costs fail even to be single-dipped: there is a minimum investment size $I > 0$, meaning that investing $i \in (0, I)$ costs $C(i) = \infty$ (whereas investing $i \geq I$ has finite cost). Our fundamental result, Theorem 1, can accommodate such failures of single-dippedness: it remains true that a rise in the marginal profitability of capital increases investment, just as would be the case if adjustment were costless. Our remaining results (Theorems 2, 3 and 4) cannot be applied in this case, however.

All of these results generalize to multiple factors of production, on the pattern of section 4.1. It suffices to assume that the factors $x = (x_1, \dots, x_n)$ are complements in production, meaning that the production function $f(x, \eta)$ is supermodular in x . Then, denoting factor prices by $r = (r_1, \dots, r_n)$, the profit function $F(x, (p, \eta, -r)) = pf(x, \eta) - r \cdot x$ is supermodular in x , and has increasing differences in $(x, (p, \eta, -r))$ as before, so that all of our general results remain applicable. In case there are just $n = 2$ factors of production, the complementarity hypothesis may be replaced with substitutability (submodularity

of $f(x, \eta)$ in x), using the trick described in section 4.1.

6 Dynamic adjustment by short-lived agents

In this section, we continue our study of the dynamic adjustment model introduced in the previous section, under a different behavioral assumption: that each period's decision is made by a short-lived agent. We recover the le Chatelier principle, and provide conditions under which short-lived agents adjust more sluggishly than a long-lived agent would.

Recall the model from section 5.1. We now assume that there is one agent per period. The period- t agent takes her predecessor's choice $x_{t-1} \in L$ as given, and chooses $x_t \in L$ to maximize the period- t payoff $G_t(\cdot, x_{t-1})$, where $G_t(x, y) = F(x, \theta_t) - C_t(x - y)$. A (*short-lived*) *equilibrium sequence* is a sequence $(x_t)_{t=1}^\infty$ in L such that $x_t \in \arg \max_{x \in L} G_t(x, x_{t-1})$ for every $t \in \mathbb{N}$, where (as before) $x_0 = \underline{x} \in \arg \max_{x \in L} F(x, \underline{\theta})$ is given.

Theorem 5 (short-lived dynamic le Chatelier). *Suppose that the objective $F(x, \theta)$ is quasi-supermodular in x and has single-crossing differences in (x, θ) , and that each adjustment cost C_t is monotone. Assume that $\arg \max_{x \in L} G_t(x, y)$ is nonempty for all $t \in \mathbb{N}$ and $y \in L$. Fix $\bar{\theta} > \underline{\theta}$, and let $\bar{x} \in \arg \max_{x \in L} F(x, \bar{\theta})$ satisfy $\bar{x} \geq \underline{x}$.¹⁸*

- If $\underline{\theta} \leq \theta_t \leq \bar{\theta}$ for every $t \in \mathbb{N}$, then there is an equilibrium sequence $(x_t)_{t=1}^\infty$ that satisfies $\underline{x} \leq x_t \leq \bar{x}$ for every $t \in \mathbb{N}$.
- If $\underline{\theta} \leq \theta_t \leq \theta_{t+1} \leq \bar{\theta}$ for every $t \in \mathbb{N}$, then there is an equilibrium sequence $(x_t)_{t=1}^\infty$ that satisfies $\underline{x} \leq x_t \leq x_{t+1} \leq \bar{x}$ for every $t \in \mathbb{N}$.

By inspection, Proposition 2 (section 4) is the special case of Theorem 5 in which there are only two periods of adjustment ($C_t(\epsilon) = \infty$ for all $t \geq 3$ and $\epsilon \neq 0$) and the parameter shifts once and for all ($\theta_1 = \theta_2 = \bar{\theta}$).

Taken together, our three dynamic le Chatelier principles (Theorems 3–5) tell us that long- and short-lived agents adjust in the same direction. The *speed* of adjustment generally differs, however. The following result gives additional assumptions under which short-lived agents adjust more sluggishly.

¹⁸Such an \bar{x} must exist, provided the argmax is nonempty (refer to footnote 10).

Theorem 6 (short- vs. long-lived). *Suppose that the objective $F(x, \theta)$ is supermodular in x and has single-crossing differences in (x, θ) , and that each adjustment cost C_t is monotone, additively separable and convex. Let $\{F(\cdot, \theta)\}_{\theta \in \Theta}$ and $\{C_t\}_{t \in \mathbb{N}}$ be equi-BCS.¹⁹ Fix $\bar{\theta} > \underline{\theta}$, let $\bar{x} \in \arg \max_{x \in L} F(x, \bar{\theta})$ satisfy $\bar{x} \geq \underline{x}$,²⁰ and assume that $\underline{\theta} \leq \theta_t \leq \theta_{t+1} \leq \bar{\theta}$ for every $t \in \mathbb{N}$. Fix a short-lived equilibrium sequence $(\tilde{x}_t)_{t=1}^\infty$ that satisfies $\underline{x} \leq \tilde{x}_t \leq \tilde{x}_{t+1} \leq \bar{x}$ for every $t \in \mathbb{N}$.²¹ Then provided the long-lived agent's problem admits a solution, there is a solution $(x_t)_{t=1}^\infty$ that satisfies $\tilde{x}_t \leq x_t \leq \bar{x}$ for every period $t \in \mathbb{N}$.*

The intuition is straightforward: short-lived agents adjust more sluggishly because they do not take into account that the less they adjust today, the more adjustment will be required tomorrow. Note that this result relies on stronger assumptions than our earlier results: in particular, convexity of costs.

The proof of Theorem 6 (appendix I) establishes that if $(x_t)_{t=1}^\infty$ solves the long-lived agent's problem and satisfies $\underline{x} \leq x_t \leq \bar{x}$ for every $t \in \mathbb{N}$,²² then $(\tilde{x}_t \vee x_t)_{t=1}^\infty$ also solves the long-lived agent's problem. By equi-BCS and a limit argument, it suffices to show that for every $T \in \{0, 1, 2, \dots\}$, the sequence $(\tilde{x}_{\min\{t, T\}} \vee x_t)_{t=1}^\infty$ solves the long-lived agent's problem, where $\tilde{x}_0 = \underline{x}$. We show this by induction on $T \in \{0, 1, 2, \dots\}$; the base case $T = 0$ is immediate, and the induction step uses the supermodularity of the objective $F(\cdot, \theta)$ and the monotonicity, additive separability and convexity of each cost C_t .

Appendix

A Standard definitions

Given $X \subseteq \mathbb{R}$, a function $\psi : X \rightarrow (-\infty, \infty]$ is *single-dipped* exactly if there is an $x \in X$ such that ψ is decreasing on $\{y \in X : y \leq x\}$ and increasing on $\{y \in X : y \geq x\}$. Given $X \subseteq \mathbb{R}^n$, a function $\psi : X \rightarrow (-\infty, \infty]$ is *bounded on compact sets (BCS)* exactly if for each compact $Y \subseteq X$, there is a constant $K_Y > 0$ such that $|\psi(y)| \leq K_Y$ for every $y \in Y$.

¹⁹Given $X \subseteq \mathbb{R}^n$, a collection $\{\phi_k\}_{k \in \mathcal{K}}$ of functions $\phi_k : X \rightarrow (-\infty, \infty]$ is *equi-BCS* if and only if the map $x \mapsto \sup_{k \in \mathcal{K}} |\phi_k(x)|$ is BCS.

²⁰Such an \bar{x} must exist, provided the argmax is nonempty (refer to footnote 10).

²¹Such a sequence exists by Theorem 5, provided there is an equilibrium sequence.

²²Such a solution exists by Theorem 3, provided there is a solution.

Fix a sublattice L of \mathbb{R}^n . A function $\phi : L \rightarrow \mathbb{R}$ is called *supermodular* if $\phi(x) - \phi(x \wedge y) \leq \phi(x \vee y) - \phi(y)$ for any $x, y \in L$, *quasi-supermodular* if $\phi(x) - \phi(x \wedge y) \geq (>) 0$ implies $\phi(x \vee y) - \phi(y) \geq (>) 0$, and *(quasi-)submodular* if $-\phi$ is (quasi-)supermodular. Clearly supermodularity implies quasi-supermodularity. If $n = 1$, then every function $\phi : L \rightarrow \mathbb{R}$ is automatically supermodular.

Fix a partially ordered set Θ . A function $F : L \times \Theta \rightarrow \mathbb{R}$ has *(strict) increasing differences* if $F(y, \theta) - F(x, \theta)$ is (strictly) increasing in θ whenever $x \leq y$, has *single-crossing differences* if $F(y, \theta') - F(x, \theta') \geq (>) 0$ implies $F(y, \theta'') - F(x, \theta'') \geq (>) 0$ whenever $x \leq y$ and $\theta' \leq \theta''$, has *strict single-crossing differences* if $F(y, \theta') - F(x, \theta') \geq 0$ implies $F(y, \theta'') - F(x, \theta'') > 0$ whenever $x < y$ and $\theta' < \theta''$, and has *(strict) decreasing differences* if $-F$ has (strict) increasing differences. A function $F : L \times \Theta \rightarrow \mathbb{R}_{++}$ has *(strict) log increasing differences* exactly if $\ln F$ has (strict) increasing differences. (Strict) increasing differences and (strict) log increasing differences each imply (strict) single-crossing differences.

Quasi-supermodularity and single-crossing differences are *ordinal* properties: they are preserved by strictly increasing transformations.²³ By contrast, supermodularity and increasing differences are in general preserved only by strictly increasing *affine* transformations: in other words, they are *cardinal*.

The sum of quasi-supermodular functions need not be quasi-supermodular. Likewise, single-crossing differences is not preserved by summation. By contrast, the sum of supermodular functions is supermodular, and the sum of functions with increasing differences also has increasing differences.

B Extension: uncertain adjustment cost

Several of our results are robust to uncertainty about adjustment costs. For Theorems 1 and 2, augment the setting from section 2 as follows. Let all uncertainty be summarized by a random variable S , called “the state of the world.” The agent’s adjustment cost is $C_s(\cdot)$ in state $S = s$. Her ex-ante payoff is

$$\tilde{G}(x, \theta) = \mathbb{E}[u(F(x, \theta) - C_S(x - \underline{x}))],$$

²³That is, for any strictly increasing $f : \mathbb{R} \rightarrow \mathbb{R}$, if ϕ is quasi-supermodular, then so is $f \circ \phi$, and if $F(x, \theta)$ has increasing differences in (x, θ) , then so does $\tilde{F}(x, \theta) = f(F(x, \theta))$.

where u is an increasing function $\mathbb{R} \rightarrow \mathbb{R}$. The curvature of the utility function u captures the agent's risk attitude.

Theorem 1'. *Suppose that the objective $F(x, \theta)$ is quasi-supermodular in x and has single-crossing differences in (x, θ) , and that at almost every realization s of the state S , the adjustment cost C_s is minimally monotone. If $\bar{\theta} > \underline{\theta}$, then $\hat{x} \geq \underline{x}$ for some $\hat{x} \in \arg \max_{x \in L} \tilde{G}(x, \bar{\theta})$, provided the argmax is nonempty.*

Proof. Let $x' \in \arg \max_{x \in L} \tilde{G}(x, \bar{\theta})$, and let $\hat{x} = \underline{x} \vee x'$. For almost every realization s , applying the proof of Theorem 1 yields $F(\hat{x}, \bar{\theta}) - C_s(\hat{x} - \underline{x}) \geq F(x', \bar{\theta}) - C_s(x' - \underline{x})$. Hence $\tilde{G}(\hat{x}, \bar{\theta}) \geq \tilde{G}(x', \bar{\theta})$ as u is increasing, which since x' maximizes $\tilde{G}(\cdot, \bar{\theta})$ on L implies that \hat{x} does, too. Clearly $\hat{x} \geq \underline{x}$. **QED**

Theorem 2'. *Suppose that the objective $F(x, \theta)$ is quasi-supermodular in x and has single-crossing differences in (x, θ) , and that at almost every realization s of the state S , the adjustment cost C_s is monotone. Fix $\bar{\theta} > \underline{\theta}$, and let $\bar{x} \in \arg \max_{x \in L} F(x, \bar{\theta})$ satisfy $\bar{x} \geq \underline{x}$.²⁴ Then*

- $\bar{x} \geq \hat{x} \geq \underline{x}$ for some $\hat{x} \in \arg \max_{x \in L} \tilde{G}(x, \bar{\theta})$, provided the argmax is nonempty, and
- if \bar{x} is the largest element of $\arg \max_{x \in L} F(x, \bar{\theta})$, then $\bar{x} \geq \hat{x}$ for any $\hat{x} \in \arg \max_{x \in L} \tilde{G}(x, \bar{\theta})$.

The proof is that of Theorem 2, modified along the lines of the proof of Theorem 1' (above). We omit the details.

To extend our dynamic results, augment the dynamic model of section 5.1 as follows. Let the period- t adjustment cost function be C_{S_t} , where S_t is a random variable. The agent's period- t expected payoff is $\tilde{G}_t(x_t, x_{t-1}) = \mathbb{E}[u(F(x_t, \theta_t) - C_{S_t}(x_t - x_{t-1}))]$, where $u : \mathbb{R} \rightarrow \mathbb{R}$ is increasing. Given her period-0 choice $x_0 \in L$, the long-lived agent's problem is to choose a sequence $(x_t)_{t=1}^\infty$ in L to maximize $\tilde{\mathcal{G}}((x_t)_{t=1}^\infty, x_0) = \sum_{t=1}^\infty \delta^{t-1} \tilde{G}_t(x_t, x_{t-1})$. We assume that the states $(S_t)_{t=1}^\infty$ are independent across periods.²⁵

²⁴Such an \bar{x} must exist, provided the argmax is nonempty (refer to footnote 10).

²⁵Theorem 3' is true without independence, but our *interpretation* of it hinges on independence. The result is about sequences $(x_t)_{t=1}^\infty$ that maximize the time-0 expected payoff $\tilde{\mathcal{G}}(\cdot, x_0)$, not about optimal real-time choice of actions. The time-0 and real-time problems are equivalent if and only if the agent does not learn over time about the realizations of future states, and this is ensured by independence.

Theorem 3'. *Suppose that the objective $F(x, \theta)$ is quasi-supermodular in x and has single-crossing differences in (x, θ) , and that in each period t , at almost every realization s_t of the state S_t , the adjustment cost C_{s_t} is monotone. Fix $\bar{\theta} > \underline{\theta}$, and let $\bar{x} \in \arg \max_{x \in L} F(x, \bar{\theta})$ satisfy $\bar{x} \geq \underline{x}$.²⁶ If $\underline{\theta} \leq \theta_t \leq \bar{\theta}$ for every $t \in \mathbb{N}$, then provided the long-lived agent's problem admits a solution, there is a solution $(x_t)_{t=1}^\infty$ that satisfies $\underline{x} \leq x_t \leq \bar{x}$ for every period $t \in \mathbb{N}$.*

Again, a simple “a.s.” modification to the proof Theorem 3 delivers this result. By contrast, the proof of Theorem 4 is not easily modified to accommodate uncertain cost, except in the risk-neutral case (when u is affine).

C Two monotonicity lemmata

The following lemma will be used in several of our appendix proofs.

Lemma 1. *Let $C : \Delta L \rightarrow \mathbb{R}$ be monotone, and consider $x, y, z \in \Delta L$.*

- *If $z \geq x$, then $C(z \wedge y - x) \leq C(y - x)$.*
- *If $z \leq x$, then $C(z \vee y - x) \leq C(y - x)$.*

We implicitly proved the first part of this lemma in the text, as part of the proof of the second part of Theorem 2.

Proof. For the first part, suppose that $z \geq x$. For each dimension i , we have either $y_i \geq z_i$ or $y_i < z_i$. If the former, then $0 \leq (y \wedge z - x)_i \leq (y - x)_i$ since $x \leq z$; if the latter, then $(y \wedge z - x)_i = (y - x)_i$. Every dimension i therefore satisfies either $0 \leq (y \wedge z - x)_i \leq (y - x)_i$ or $0 \geq (y \wedge z - x)_i \geq (y - x)_i$, which by monotonicity implies that $C(y \wedge z - x) \leq C(y - x)$.

Similarly, for the second part, suppose that $z \leq x$. Each dimension i has either $y_i \leq z_i$ or $y_i > z_i$. If the former, then $0 \geq (y \vee z - x)_i \geq (y - x)_i$ by $x \geq z$; if the latter, then $(y \vee z - x)_i = (y - x)_i$. Thus $C(y \vee z - x) \leq C(y - x)$ by monotonicity. **QED**

The following variant will be also be used occasionally.

Lemma 2. *If $C : \Delta L \rightarrow \mathbb{R}$ is monotone, then $C(z \vee x - z \vee y) \leq C(x - y) \geq C(z \wedge x - z \wedge y)$ for any $x, y, z \in \Delta L$.*

²⁶Such an \bar{x} must exist, provided the argmax is nonempty (refer to footnote 10).

Proof. We shall prove the first inequality; the second follows similarly. By monotonicity, it suffices to show that for each i , one of the following holds:

$$(a) \ 0 \leq (z \vee x - z \vee y)_i \leq (x - y)_i$$

$$(b) \ 0 \geq (z \vee x - z \vee y)_i \geq (x - y)_i.$$

If $x_i \geq z_i \geq y_i$ then (a) holds by inspection, if $x_i \leq z_i \leq y_i$ then (b) holds by inspection, if $x_i \leq z_i \geq y_i$ then (a) or (b) holds since $(z \vee x - z \vee y)_i = 0$, and if $x_i \geq z_i \leq y_i$ then both (a) and (b) hold since $(z \vee x - z \vee y)_i = (x - y)_i$. **QED**

D Proof of Proposition 1

Let $\hat{x} \in \arg \max_{x \in L} G(x, \bar{\theta})$, and suppose toward a contradiction that $\hat{x} \not\geq \underline{x}$. Then $\underline{x} \vee \hat{x} > \hat{x}$. The proof of Theorem 1 yields $F(\underline{x} \vee \hat{x}, \bar{\theta}) \geq F(\hat{x}, \bar{\theta})$ and $C(\underline{x} \vee \hat{x} - \underline{x}) \leq C(\hat{x} - \underline{x})$, where the first inequality is strict if the single-crossing differences of F is strict, and the second inequality is strict if the minimal monotonicity of C is strict. In either case, we have

$$G(\underline{x} \vee \hat{x}, \bar{\theta}) = F(\underline{x} \vee \hat{x}, \bar{\theta}) - C(\underline{x} \vee \hat{x} - \underline{x}) > F(\hat{x}, \bar{\theta}) - C(\hat{x} - \underline{x}) = G(\hat{x}, \bar{\theta}),$$

which contradicts the fact that \hat{x} maximizes $G(\cdot, \bar{\theta})$ on L . **QED**

E Proof of Theorem 3

Let $(x_t)_{t=1}^\infty$ maximize $\mathcal{G}(\cdot, x_0)$. We shall show that $(\bar{x} \wedge (\underline{x} \vee x_t))_{t=1}^\infty$ also maximizes $\mathcal{G}(\cdot, x_0)$; this suffices since $\underline{x} \leq \bar{x} \wedge (\underline{x} \vee x_t) \leq \bar{x}$ for each t .

We first show that $(\hat{x}_t)_{t=1}^\infty = (\underline{x} \vee x_t)_{t=1}^\infty$ maximizes $\mathcal{G}(\cdot, x_0)$. For every $t \in \mathbb{N}$, we have $F(\underline{x}, \underline{\theta}) \geq F(\underline{x} \wedge x_t, \underline{\theta})$ by definition of \underline{x} , which by quasi-supermodularity implies that $F(\underline{x} \vee x_t, \underline{\theta}) \geq F(x_t, \underline{\theta})$, whence $F(\underline{x} \vee x_t, \theta_t) \geq F(x_t, \theta_t)$ by single-crossing differences and $\theta_t \geq \underline{\theta}$. Thus $\mathcal{F}((\underline{x} \vee x_t)_{t=1}^\infty) \geq \mathcal{F}((x_t)_{t=1}^\infty)$. Furthermore, for every $t \in \mathbb{N}$, we have $C_t(\underline{x} \vee x_t - \underline{x} \vee x_{t-1}) \leq C_t(x_t - x_{t-1})$ by Lemma 2 (appendix C) since C_t is monotone, so $\mathcal{C}(x_0, (\underline{x} \vee x_t)_{t=1}^\infty) \leq \mathcal{C}(x_0, (x_t)_{t=1}^\infty)$. So $\mathcal{G}((\underline{x} \vee x_t)_{t=1}^\infty, x_0) \geq \mathcal{G}((x_t)_{t=1}^\infty, x_0)$, which since $(x_t)_{t=1}^\infty$ maximizes $\mathcal{G}(\cdot, x_0)$ implies that $(\hat{x}_t)_{t=1}^\infty = (\underline{x} \vee x_t)_{t=1}^\infty$ does, too.

It remains to show that $(\bar{x} \wedge \hat{x}_t)_{t=1}^\infty$ also maximizes $\mathcal{G}(\cdot, x_0)$. For every $t \in \mathbb{N}$, we have $F(\bar{x} \vee \hat{x}_t, \bar{\theta}) \leq F(\bar{x}, \bar{\theta})$ by definition of \bar{x} , which by quasi-

supermodularity implies that $F(\hat{x}_t, \bar{\theta}) \leq F(\bar{x} \wedge \hat{x}_t, \bar{\theta})$, whence $F(\hat{x}_t, \theta_t) \leq F(\bar{x} \wedge \hat{x}_t, \theta_t)$ by single-crossing differences and $\theta_t \leq \bar{\theta}$. Thus $\mathcal{F}((\hat{x}_t)_{t=1}^\infty) \leq \mathcal{F}((\bar{x} \wedge \hat{x}_t)_{t=1}^\infty)$. Furthermore, for every $t \in \mathbb{N}$, we have $C_t(\hat{x}_t - \hat{x}_{t-1}) \geq C_t(\bar{x} \wedge \hat{x}_t - \bar{x} \wedge \hat{x}_{t-1})$ by Lemma 2 since C_t is monotone, so $\mathcal{C}(x_0, (\hat{x}_t)_{t=1}^\infty) \geq \mathcal{C}(x_0, (\bar{x} \wedge \hat{x}_t)_{t=1}^\infty)$. So $\mathcal{G}((\hat{x}_t)_{t=1}^\infty, x_0) \leq \mathcal{G}((\bar{x} \wedge \hat{x}_t)_{t=1}^\infty, x_0)$, which since $(\hat{x}_t)_{t=1}^\infty$ maximizes $\mathcal{G}(\cdot, x_0)$ implies that $(\bar{x} \wedge \hat{x}_t)_{t=1}^\infty$ does, too. **QED**

F Proof of Theorem 4

For any sequence $\mathbf{x} = (x_t)_{t=1}^\infty$ in L and any $T \in \mathbb{N}$, let $M_T \mathbf{x}$ denote the sequence in L whose t^{th} entry is x_t for $t < T$ and $x_{t-1} \vee x_t$ for $t \geq T$.

Assume that the agent's problem admits a solution. Let $\mathbf{x}^1 = (x_t^1)_{t=1}^\infty$ be a solution satisfying $\underline{x} \leq x_t^1 \leq \bar{x}$ in every period t ; such a solution exists by Theorem 3. Define $X_t = x_1^1 \vee x_2^1 \vee \dots \vee x_{t-1}^1 \vee x_t^1$ for $t \in \mathbb{N}$, and $X_0 = \underline{x}$.

Write $\mathbf{x}^T = M_T M_{T-1} \dots M_3 M_2 \mathbf{x}^1$ for $T \geq 2$. By inspection, the first T entries of \mathbf{x}^T are $X_1, X_2, \dots, X_{T-1}, X_T$. Clearly $\underline{x} \leq X_t \leq X_{t+1} \leq \bar{x}$ for any period $t \in \mathbb{N}$. To prove the theorem, we need only show that $\mathbf{x}^\infty = (X_1, X_2, X_3, \dots)$ is optimal.

It suffices to show for each $T \in \mathbb{N}$ that \mathbf{x}^T is optimal. For then, letting V be the optimal value and noting that both $\mathbf{x}^T = (x_t)_{t=1}^\infty$ and \mathbf{x}^∞ have X_1, \dots, X_T as their first T entries, we have

$$\begin{aligned}
0 &\geq \mathcal{G}(\mathbf{x}^\infty, x_0) - V = \mathcal{G}(\mathbf{x}^\infty, x_0) - \mathcal{G}(\mathbf{x}^T, x_0) \\
&= \delta^T [\mathcal{G}((X_t)_{t=T+1}^\infty, X_T) - \mathcal{G}((x_t)_{t=T+1}^\infty, X_T)] \\
&= \delta^T [\mathcal{F}((X_t)_{t=T+1}^\infty) - \mathcal{F}((x_t)_{t=T+1}^\infty)] \\
&\quad - \delta^T [\mathcal{C}(X_T, (X_t)_{t=T+1}^\infty) - \mathcal{C}(X_T, (x_t)_{t=T+1}^\infty)] \\
&\geq \delta^T [\mathcal{F}((X_t)_{t=T+1}^\infty) - \mathcal{F}((x_t)_{t=T+1}^\infty)], \tag{1}
\end{aligned}$$

where the final inequality holds since

$$\begin{aligned}
&C(X_t - X_{t-1}) \leq C(x_t - x_{t-1}) \quad \text{for every } t \geq T+2 \\
&\text{and } C(X_{T+1} - X_T) \leq C(x_{T+1} - X_T)
\end{aligned}$$

by the monotonicity of C . (For $t \geq T+2$, for each dimension i , if $X_{t,i} = X_{t-1,i}$

then $0 = (X_t - X_{t-1})_i$, while if $X_{t,i} > X_{t-1,i}$ then $x_{t,i} = X_{t,i} \geq X_{t-1,i} \geq x_{t-1,i}$, so $0 \leq (X_t - X_{t-1})_i \leq (x_t - x_{t-1})_i$. For the $t = T + 1$ inequality, if $X_{T+1,i} = X_{T,i}$ then $0 = (X_{T+1} - X_T)_i \geq (x_{T+1} - X_T)_i$, while if $X_{T+1,i} > X_{T,i}$ then $X_{T+1,i} = x_{T+1,i}$, so $(X_{T+1} - X_T)_i = (x_{T+1} - X_T)_i$. Since $F(\cdot, \bar{\theta})$ is BCS, the “[.]” expression in (1) is bounded below uniformly over $T \in \mathbb{N}$,²⁷ so letting $T \rightarrow \infty$ yields $0 \geq \mathcal{G}(\mathbf{x}^\infty, x_0) - V \geq 0$, which is to say that \mathbf{x}^∞ is optimal.

To show that \mathbf{x}^T is optimal for each $T \in \mathbb{N}$, we employ induction on $T \in \mathbb{N}$. The base case $T = 1$ is immediate.

For the induction step, fix any $T \in \mathbb{N}$, and suppose that $\mathbf{x}^T = (x_t)_{t=1}^\infty$ is optimal; we will show that $\mathbf{x}^{T+1} = M_{T+1}\mathbf{x}^T$ is also optimal. Let $(\tilde{x}_t)_{t=1}^\infty$ be the sequence with t^{th} entry x_t for $t < T$ and $x_t \wedge x_{t+1}$ for $t \geq T$. Since $\mathbf{x}^T = (x_t)_{t=1}^\infty$ is optimal, and $(\tilde{x}_t)_{t=1}^\infty$ shares its first $T - 1$ entries X_1, \dots, X_{T-1} , we have $\mathcal{G}((x_t)_{t=T}^\infty, X_{T-1}) \geq \mathcal{G}((\tilde{x}_t)_{t=T}^\infty, X_{T-1})$, which may be written in full as

$$\begin{aligned} & \sum_{t=T}^{\infty} \delta^{t-T} [F(x_t, \bar{\theta}) - F(x_t \wedge x_{t+1}, \bar{\theta})] \\ & - \sum_{t=T}^{\infty} \delta^{t-T} [C(x_t - x_{t-1}) - C(x_t \wedge x_{t+1} - x_{t-1} \wedge x_t)] \geq 0. \end{aligned} \quad (2)$$

(Note that since $x_t = X_t$ for every $t \leq T$, we have $x_{T-1} \wedge x_T = X_{T-1} = x_{T-1}$.) Since $F(\cdot, \bar{\theta})$ is supermodular, it holds for every $t \geq T$ that

$$F(x_t \vee x_{t+1}, \bar{\theta}) - F(x_{t+1}, \bar{\theta}) \geq F(x_t, \bar{\theta}) - F(x_t \wedge x_{t+1}, \bar{\theta}) \quad (3)$$

We furthermore claim that for each $t \geq T$,

$$\begin{aligned} & C(x_t \vee x_{t+1} - x_{t-1} \vee x_t) - C(x_{t+1} - x_t) \\ & \leq C(x_t - x_{t-1}) - C(x_t \wedge x_{t+1} - x_{t-1} \wedge x_t); \end{aligned} \quad (4)$$

we shall prove this shortly. Combining (2), (3) and (4), and changing variables

²⁷Since X_t and x_t belong to the compact set $[\underline{x}, \bar{x}]$ for every $t \in \mathbb{N}$, there is a $K > 0$ such that $F(X_t, \bar{\theta}) - F(x_t, \bar{\theta}) \geq -2K$ for all t , so “[.]” is bounded below by $-2K/(1 - \delta)$.

in the sums, we obtain

$$\begin{aligned} & \sum_{t=T+1}^{\infty} \delta^{t-(T+1)} [F(x_{t-1} \vee x_t, \bar{\theta}) - F(x_t, \bar{\theta})] \\ & - \sum_{t=T+1}^{\infty} \delta^{t-(T+1)} [C(x_{t-1} \vee x_t - x_{t-2} \vee x_{t-1}) - C(x_t - x_{t-1})] \geq 0. \end{aligned}$$

By inspection, this says precisely that $(\hat{x}_t)_{t=1}^{\infty} = \mathbf{x}^{T+1} = M_{T+1} \mathbf{x}^T$ satisfies

$$\mathcal{G}((\hat{x}_t)_{t=T+1}^{\infty}, X_T) \geq \mathcal{G}((x_t)_{t=T+1}^{\infty}, X_T).$$

(Note that since $x_t = X_t$ for every $t \leq T$, we have $x_{(T+1)-2} \vee x_{(T+1)-1} = X_{T-1} \vee X_T = X_T = x_{(T+1)-1}$.) Since $\mathbf{x}^{T+1} = (\hat{x}_t)_{t=1}^{\infty}$ and $\mathbf{x}^T = (x_t)_{t=1}^{\infty}$ agree in their first T entries, and \mathbf{x}^T is optimal, it follows that \mathbf{x}^{T+1} is optimal, too.

It remains to show that (4) holds. It suffices to prove for each i that

$$C_i(y \vee z - x \vee y) + C_i(y \wedge z - x \wedge y) \leq C_i(y - x) + C_i(z - y) \quad \text{for any } x, y, z. \quad (5)$$

(We've renamed $x_{t-1,i} = x$, $x_{t,i} = y$ and $x_{t+1,i} = z$.) When y is not extreme (neither least nor greatest), (5) holds trivially because the left-hand side is equal to the right-hand side. When y is extreme, (5) reads

$$C_i(z - x) \leq C_i(y - x) + C_i(z - y),$$

and we have either

$$\begin{aligned} & \text{(i) } 0 \leq z - x \leq y - x \quad \text{or} \quad \text{(ii) } 0 \leq z - x \leq z - y \\ \text{or} \quad & \text{(iii) } 0 \geq z - x \geq y - x \quad \text{or} \quad \text{(iv) } 0 \geq z - x \geq z - y. \end{aligned}$$

Since C_i is single-dipped, minimized at zero and nonnegative, we have $C_i(z - x) \leq C(y - x) \leq C(y - x) + C(z - y)$ in the first and third cases, and $C_i(z - x) \leq C(z - y) \leq C(y - x) + C(z - y)$ in the second and fourth. **QED**

G Proof of Theorem 5

For the first part, fix a sequence $(\theta_t)_{t=1}^\infty$ in Θ such that $\underline{\theta} \leq \theta_t \leq \bar{\theta}$ for every $t \in \mathbb{N}$. Call a finite sequence $(x_t)_{t=1}^T$ *equilibrium caged* exactly if $x_t \in \arg \max_{x \in L} G_t(x, x_{t-1})$ and $\underline{x} \leq x_t \leq \bar{x}$ for every $t \in \{1, \dots, T\}$. By Theorem 2, there exists an equilibrium caged sequence of length $T = 1$. Given this, it suffices to prove that for each $T \geq 2$, any length- $(T - 1)$ equilibrium caged sequence $(x_t)_{t=1}^{T-1}$ may be extended to a length- T equilibrium caged sequence $(x_t)_{t=1}^T$ (by appropriately choosing x_T). To that end, fix an arbitrary $T \geq 2$, and let $(x_t)_{t=1}^{T-1}$ be equilibrium caged. We shall prove two claims:

Claim 1. *There is an $x' \in \arg \max_{x \in L} G_T(x, x_{T-1})$ such that $x' \geq \underline{x}$.*

Claim 2. *$\bar{x} \wedge x'$ belongs to $\arg \max_{x \in L} G_T(x, x_{T-1})$ whenever x' does.*

These claims suffice because $x_T = \bar{x} \wedge x'$ satisfies $\underline{x} \leq x_T \leq \bar{x}$.

Proof of Claim 1. Fix any $x'' \in \arg \max_{x \in L} G_T(x, x_{T-1})$. We will show that $x' = \underline{x} \vee x''$ also maximizes $G_T(\cdot, x_{T-1})$; obviously $x' \geq \underline{x}$. We have $F(\underline{x}, \underline{\theta}) \geq F(\underline{x} \wedge x'', \underline{\theta})$ by definition of \underline{x} . Thus $F(\underline{x} \vee x'', \underline{\theta}) \geq F(x'', \underline{\theta})$ by quasi-supermodularity, whence $F(\underline{x} \vee x'', \theta_T) \geq F(x'', \theta_T)$ by single-crossing differences and $\theta_T \geq \underline{\theta}$. Furthermore, since C_T is monotone and $\underline{x} \leq x_{T-1}$, we have $C_T(\underline{x} \vee x' - x_{T-1}) \leq C_T(x' - x_{T-1})$ by Lemma 1. Thus

$$\begin{aligned} G_T(x', x_{T-1}) &= F(\underline{x} \vee x'', \theta_T) - C_T(\underline{x} \vee x'' - x_{T-1}) \\ &\geq F(x'', \theta_T) - C_T(x'' - x_{T-1}) = G_T(x'', x_{T-1}). \end{aligned}$$

Since x'' maximizes $G_T(\cdot, x_{T-1})$ on L , it follows that x' does, too. QED

Proof of Claim 2. Let x' belong to $\arg \max_{x \in L} G_T(x, x_{T-1})$; we claim that $\hat{x} = \bar{x} \wedge x'$ also maximizes $G_T(\cdot, x_{T-1})$. We have $F(\bar{x} \vee x', \bar{\theta}) \leq F(\bar{x}, \bar{\theta})$ by definition of \bar{x} , whence $F(x', \bar{\theta}) \leq F(\bar{x} \wedge x', \bar{\theta})$ by quasi-supermodularity, so that $F(x', \theta_T) \leq F(\bar{x} \wedge x', \theta_T)$ by single-crossing differences and $\theta_T \leq \bar{\theta}$. Since C_T is monotone and $\bar{x} \geq x_{T-1}$, we have $C_T(x' - x_{T-1}) \geq C_T(\bar{x} \wedge x' - x_{T-1})$ by Lemma 1. Thus

$$\begin{aligned} G_T(x', x_{T-1}) &= F(x', \theta_T) - C_T(x' - x_{T-1}) \\ &\leq F(\bar{x} \wedge x', \theta_T) - C_T(\bar{x} \wedge x' - x_{T-1}) = G_T(\hat{x}, x_{T-1}), \end{aligned}$$

which since x' maximizes $G_T(\cdot, x_{T-1})$ on L implies that \hat{x} does, too. *QED*

To prove the second part of Theorem 5, fix a sequence $(\theta_t)_{t=1}^\infty$ in Θ such that $\underline{\theta} \leq \theta_t \leq \theta_{t+1} \leq \bar{\theta}$ for every $t \in \mathbb{N}$. Recall that $x_0 = \underline{x}$. Call a finite sequence $(x_t)_{t=1}^T$ *equilibrium monotone* exactly if $x_t \in \arg \max_{x \in L} G_t(x, x_{t-1})$ and $x_{t-1} \leq x_t \leq \bar{x}$ for every $t \in \{1, \dots, T\}$. By Theorem 2, there exists an equilibrium monotone sequence of length $T = 1$. Given this, it suffices to prove that for every $T \geq 2$, any length- $(T-1)$ equilibrium monotone sequence $(x_t)_{t=1}^{T-1}$ may be extended to a length- T equilibrium monotone sequence $(x_t)_{t=1}^T$ (by an appropriate choice of x_T).

To that end, fix an arbitrary $T \geq 2$, and let $(x_t)_{t=1}^{T-1}$ be equilibrium monotone; we shall show that for every $t \in \{1, \dots, T\}$, there is an $x' \in L$ which belongs to $\arg \max_{x \in L} G_T(x, x_{T-1})$ and satisfies $x_{t-1} \leq x' \leq \bar{x}$. We proceed by induction on $t \in \{1, \dots, T\}$. The base case $t = 1$ follows from Claims 1 and 2. For the induction step, suppose that there is an $x'' \in \arg \max_{x \in L} G_T(x, x_{T-1})$ that satisfies $x_{t-2} \leq x'' \leq \bar{x}$; we claim that $x' = x_{t-1} \vee x''$ also maximizes $G_T(\cdot, x_{T-1})$. This suffices since $x_{t-1} \leq x' \leq \bar{x}$, where the latter inequality holds because $x_{t-1} \leq \bar{x}$ (as $(x_s)_{s=1}^{T-1}$ is equilibrium monotone) and $x'' \leq \bar{x}$.

We have $G_{t-1}(x_{t-1}, x_{t-2}) \geq G_{t-1}(x_{t-1} \wedge x'', x_{t-2})$ by definition of x_{t-1} . Since C_{t-1} is monotone and $x'' \geq x_{t-2}$ by the induction hypothesis, we have $C_{t-1}(x_{t-1} - x_{t-2}) \geq C_{t-1}(x_{t-1} \wedge x'' - x_{t-2})$ by Lemma 1. It follows that $F(x_{t-1}, \theta_{t-1}) \geq F(x_{t-1} \wedge x'', \theta_{t-1})$. Thus $F(x_{t-1} \vee x'', \theta_{t-1}) \geq F(x'', \theta_{t-1})$ by quasi-supermodularity, whence $F(x_{t-1} \vee x'', \theta_T) \geq F(x'', \theta_T)$ by single-crossing differences and $\theta_T \geq \theta_{t-1}$. We have $C_T(x_{t-1} \vee x'' - x_{T-1}) \leq C_T(x'' - x_{T-1})$ by Lemma 1 since C_T is monotone and $x_{t-1} \leq x_{T-1}$, where the latter holds since $(x_s)_{s=1}^{T-1}$ is equilibrium monotone. Thus

$$\begin{aligned} G_T(x', x_{T-1}) &= F(x_{t-1} \vee x'', \theta_T) - C_T(x_{t-1} \vee x'' - x_{T-1}) \\ &\geq F(x'', \theta_T) - C_T(x'' - x_{T-1}) = G_T(x'', x_{T-1}), \end{aligned}$$

which since x'' maximizes $G_T(\cdot, x_{T-1})$ on L implies that x' does, too. *QED*

H Proof of Proposition 2

Define a sequence $(\theta_t)_{t=1}^\infty$ in Θ by $\theta_t = \bar{\theta}$ for every t . For each $t \geq 3$, define $C_t : \Delta L \rightarrow [0, \infty]$ by $C_t(\epsilon) = \infty$ for every $\epsilon \neq 0$ and $C_t(0) = 0$. Now apply Theorem 5. **QED**

I Proof of Theorem 6

For any sequence $\mathbf{x} = (x_t)_{t=1}^\infty$ in L and any $T \in \{0, 1, 2, \dots\}$, let $R_T \mathbf{x}$ denote the sequence in L whose t^{th} entry is x_t for $t < T$ and $\tilde{x}_T \vee x_t$ for $t \geq T$.

The long-lived agent's problem is to maximize $\mathcal{G}(\cdot, x_0)$. Assume that it admits a solution. Let $\mathbf{x}' = (x'_t)_{t=1}^\infty$ be a solution satisfying $\underline{x} \leq x'_t \leq \bar{x}$ in every period t ; such a solution exists by Theorem 3. Define $X_t = \tilde{x}_t \vee x'_t$ for each $t \in \mathbb{N}$, and $X_0 = \underline{x}$.

Write $\mathbf{x}^T = R_T R_{T-1} \cdots R_2 R_1 R_0 \mathbf{x}'$ for $T \in \{0, 1, 2, \dots\}$. The sequence \mathbf{x}^T has t^{th} entry X_t for $t \leq T$ and $\tilde{x}_T \vee x_t$ for $t > T$, since $(\tilde{x}_t)_{t=1}^\infty$ is increasing. Clearly $\tilde{x}_t \leq X_t \leq \bar{x}$ for any period $t \in \mathbb{N}$. To prove the theorem, we need only show that $\mathbf{x}^\infty = (X_1, X_2, X_3, \dots)$ maximizes $\mathcal{G}(\cdot, x_0)$.

It suffices to show for each $T \in \{0, 1, 2, \dots\}$ that \mathbf{x}^T maximizes $\mathcal{G}(\cdot, x_0)$. For then, letting V be the long-lived agent's optimal value and noting that both $\mathbf{x}^T = (x_t)_{t=1}^\infty$ and \mathbf{x}^∞ have X_1, \dots, X_T as their first T entries, we have

$$\begin{aligned} 0 &\geq \mathcal{G}(\mathbf{x}^\infty, x_0) - V = \mathcal{G}(\mathbf{x}^\infty, x_0) - \mathcal{G}(\mathbf{x}^T, x_0) \\ &= \delta^T [\mathcal{G}((X_t)_{t=T+1}^\infty, X_T) - \mathcal{G}((x_t)_{t=T+1}^\infty, X_T)]. \end{aligned}$$

By equi-BCS, the right-hand “[\cdot]” is bounded below uniformly over $T \in \mathbb{N}$,²⁸ so letting $T \rightarrow \infty$ yields $0 \geq \mathcal{G}(\mathbf{x}^\infty, x_0) - V \geq 0$, meaning that \mathbf{x}^∞ is optimal.

To show that \mathbf{x}^T is optimal for each $T \in \{0, 1, 2, \dots\}$, we employ induction on $T \in \{0, 1, 2, \dots\}$. The base case $T = 0$ is immediate, since $\mathbf{x}^0 = R_0 \mathbf{x}' = \mathbf{x}'$ is optimal.

For the induction step, fix any $T \in \mathbb{N}$, and suppose that $\mathbf{x}^{T-1} = (x_t)_{t=1}^\infty$ is optimal; we will show that $\mathbf{x}^T = R_T \mathbf{x}^{T-1}$ is also optimal. Since \mathbf{x}^{T-1} and \mathbf{x}^T have the same first $T-1$ entries (namely, X_1, X_2, \dots, X_{T-1}), and since for

²⁸Since X_t and x_t belong to the compact set $[\underline{x}, \bar{x}]$ for every $t \in \mathbb{N}$, there are constants $A, B > 0$ such that $F(X_t, \theta_t) - F(x_t, \theta_t) \geq -2A$ and $-[C(X_t - X_{t-1}) + C(x_t - x_{t-1})] \geq -2B$ for all $t \in \mathbb{N}$, so the right-hand “[\cdot]” is bounded below by $-2(A+B)/(1-\delta)$.

$t \geq T$ the t^{th} entry of \mathbf{x}^T is $\tilde{x}_T \vee x_t$, it suffices to show that

$$G_T(X_T, X_{T-1}) \geq G_T(x_T, X_{T-1}) \quad (6)$$

$$G_t(\tilde{x}_T \vee x_t, \tilde{x}_T \vee x_{t-1}) \geq G_t(x_t, x_{t-1}) \quad \text{for all } t \geq T+1. \quad (7)$$

For (6), since C_T is convex and $X_T \geq x_T$ and $X_{T-1} \geq \tilde{x}_{T-1}$, we have

$$C_T(X_T - X_{T-1}) - C_T(x_T - X_{T-1}) \leq C_T(X_T - \tilde{x}_{T-1}) - C_T(x_T - \tilde{x}_{T-1}).$$

It follows that

$$\begin{aligned} G_T(X_T, X_{T-1}) - G_T(x_T, X_{T-1}) &\geq G_T(X_T, \tilde{x}_{T-1}) - G_T(x_T, \tilde{x}_{T-1}) \\ &\geq G_T(\tilde{x}_T, \tilde{x}_{T-1}) - G_T(\tilde{x}_T \wedge x_T, \tilde{x}_{T-1}) \geq 0, \end{aligned}$$

where the second inequality holds since

$$\begin{aligned} F(X_T, \theta_T) - F(x_T, \theta_T) &\geq F(\tilde{x}_T, \theta_T) - F(\tilde{x}_T \wedge x_T, \theta_T) \quad \text{and} \\ C_T(X_T - \tilde{x}_{T-1}) - C_T(x_T - \tilde{x}_{T-1}) &= C_T(\tilde{x}_T - \tilde{x}_{T-1}) - C_T(\tilde{x}_T \wedge x_T - \tilde{x}_{T-1}) \end{aligned}$$

by the supermodularity of $F(\cdot, \theta_T)$ and the additive separability of C_T , and the final inequality holds since \tilde{x}_T maximizes $G_T(\cdot, \tilde{x}_{T-1})$ on L by definition.

It remains to establish (7). Fix an arbitrary $t \geq T+1$. It suffices to show that $F(\tilde{x}_T \vee x_t, \theta_t) \geq F(x_t, \theta_t)$ and $C_t(\tilde{x}_T \vee x_t - \tilde{x}_T \vee x_{t-1}) \leq C_t(x_t - x_{t-1})$. The latter holds by Lemma 2 (appendix C) since C_t is monotone. To show the former, begin by noting that $G_T(\tilde{x}_T, \tilde{x}_{T-1}) \geq G_T(\tilde{x}_T \wedge x_t, \tilde{x}_{T-1})$ since \tilde{x}_T maximizes $G_T(\cdot, \tilde{x}_{T-1})$ on L by definition. We have $C_T(\tilde{x}_T - \tilde{x}_{T-1}) \geq C_T(\tilde{x}_T \wedge x_t - \tilde{x}_{T-1})$ by Lemma 1 since $\tilde{x}_T \geq \tilde{x}_{T-1}$. It follows that $F(\tilde{x}_T, \theta_T) \geq F(\tilde{x}_T \wedge x_t, \theta_T)$. Thus $F(\tilde{x}_T \vee x_t, \theta_T) \geq F(x_t, \theta_T)$ by supermodularity, whence $F(\tilde{x}_T \vee x_t, \theta_t) \geq F(x_t, \theta_t)$ by single-crossing differences and the fact that $\theta_t \geq \theta_T$ (since $t > T$ and $(\theta_s)_{s=1}^\infty$ is increasing). **QED**

References

Antić, N., & Salant, Y. (in progress). *Monopoly theory with price-increase aversion*.

- Bachmann, R., Caballero, R. J., & Engel, E. M. R. A. (2013). Aggregate implications of lumpy investment: New evidence and a DSGE model. *American Economic Journal: Macroeconomics*, 5(4), 29–67. <https://doi.org/10.1257/mac.5.4.29>
- Bari, F., Malik, K., Meki, M., & Quinn, S. (2021). *Asset-based microfinance for microenterprises: Evidence from Pakistan* [working paper, Jan 2021]. <https://doi.org/10.2139/ssrn.3778754>
- Berger, D., & Vavra, J. (2015). Consumption dynamics during recessions. *Econometrica*, 83(1), 101–154. <https://doi.org/10.3982/ECTA11254>
- Caplin, A., & Spulber, D. F. (1987). Menu costs and the neutrality of money. *Quarterly Journal of Economics*, 102(4), 703–725. <https://doi.org/10.2307/1884277>
- Cooper, R. W., & Haltiwanger, J. C. (2006). On the nature of capital adjustment costs. *Review of Economic Studies*, 73(3), 611–633. <https://doi.org/10.1111/j.1467-937X.2006.00389.x>
- Field, E., Pande, R., Papp, J., & Rigol, N. (2013). Does the classic microfinance model discourage entrepreneurship among the poor? Experimental evidence from India. *Journal of Political Economy*, 103(6), 2196–2226. <https://doi.org/10.1257/aer.103.6.2196>
- Galí, J. (2015). *Monetary policy, inflation, and the business cycle: An introduction to the new Keynesian framework and its applications* (2nd). Princeton University Press.
- Golosov, M., & Lucas, R. E., Jr. (2007). Menu costs and Phillips curves. *Journal of Political Economy*, 115(2), 171–199. <https://doi.org/10.1086/512625>
- Grossman, S. J., & Laroque, G. (1990). Asset pricing and optimal portfolio choice in the presence of illiquid durable consumption goods. *Econometrica*, 58(1), 25–51. <https://doi.org/10.2307/2938333>
- Hayashi, F. (1982). Tobin’s marginal q and average q : A neoclassical interpretation. *Econometrica*, 50(1), 213–224. <https://doi.org/10.2307/191253>
- Hopenhayn, H. A., & Prescott, E. C. (1992). Stochastic monotonicity and stationary distributions for dynamic economies. *Econometrica*, 60(6), 1387–1406. <https://doi.org/10.2307/2951526>

- Jorgenson, D. W. (1963). Capital theory and investment behavior. *American Economic Review*, 53(2), 247–259.
- Kaplan, G., & Violante, G. L. (2014). A model of the consumption response to fiscal stimulus payments. *Econometrica*, 82(4), 1199–1239. <https://doi.org/10.3982/ECTA10528>
- Levy, D., Bergen, M., Dutta, S., & Venable, R. (1997). The magnitude of menu costs: Direct evidence from large U.S. supermarket chains. *Quarterly Journal of Economics*, 112(3), 791–824. <https://doi.org/10.1162/003355397555352>
- Mankiw, N. G. (1985). Small menu costs and large business cycles: A macroeconomic model of monopoly. *Quarterly Journal of Economics*, 100(2), 529–538. <https://doi.org/10.2307/1885395>
- Midrigan, V. (2011). Menu costs, multiproduct firms, and aggregate fluctuations. *Econometrica*, 79(4), 1139–1180. <https://doi.org/10.3982/ECTA6735>
- Milgrom, P., & Roberts, J. (1990). Rationalizability, learning, and equilibrium in games with strategic complementarities. *Econometrica*, 58(6), 1255–1277. <https://doi.org/10.2307/2938316>
- Milgrom, P., & Roberts, J. (1996). The LeChatelier principle. *American Economic Review*, 86(1), 173–179.
- Milgrom, P., & Shannon, C. (1994). Monotone comparative statics. *Econometrica*, 62(1), 157–180. <https://doi.org/10.2307/2951479>
- Rotemberg, J. J. (1982). Monopolistic price adjustment and aggregate output. *Review of Economic Studies*, 49(4), 517–531. <https://doi.org/10.2307/2297284>
- Samuelson, P. A. (1947). *Foundations of economic analysis*. Harvard University Press.
- Thomas, J. K. (2002). Is lumpy investment relevant for the business cycle? *Journal of Political Economy*, 110(3), 508–534. <https://doi.org/10.1086/339746>
- Tobin, J. (1969). A general equilibrium approach to monetary theory. *Journal of Money, Credit and Banking*, 1(1), 15–29. <https://doi.org/10.2307/1991374>

Winberry, T. (2021). Lumpy investment, business cycles, and stimulus policy.
American Economic Review, 111(1), 364–396. <https://doi.org/10.1257/aer.20161723>