CESifo WORKING PAPERS

7564
2019

March 2019

Dynamic IC and dynamic programming

Suehyun Kwon

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 7564*
Category 12: Empirical and Theoretical Methods

Dynamic IC and dynamic programming

Abstract

This paper develops a dynamic programming method when the one-stage deviation principle in the sense of mechanism design literature doesn’t hold. The commonly used dynamic programming method is valid only if the one-stage deviation principle in the sense of mechanism design literature is satisfied; it doesn't hold in every model, and the one-stage deviation principle in the sense of repeated games does hold but requires the equilibrium strategy of every player off the equilibrium path and is impractical. The dynamic programming method developed in this paper requires transfinite induction, and therefore one needs to specify the stopping times for two dimensions.

Keywords: dynamic programming, one-stage deviation, transfinite induction, stopping time.

Suehyun Kwon
Department of Economics
University College London
30 Gordon Street
United Kingdom – London, WC1H 0AX
suehyun.kwon@gmail.com

March 13, 2019
This paper includes some of the results that were circulated in “Mechanism Design with Moral Hazard.”

1 Introduction

The dynamic programming method typically used in macroeconomics is a little bit different from the contraction mapping in Abreu, Pearce and Stacchetti (1986, 1990) (APS in short). However, loosely speaking, both start with a bounded set that includes all feasible payoffs, i.e., any payoff that can be an equilibrium payoff or an expected payoff of the representative agent. Once one starts with a bounded set and applies a contraction mapping, then the image of the mapping is strictly smaller than the set the mapping is applied to; as long as there are finitely many players, there exists a unique limit. The existence of the unique limit can be shown by the dominated convergence theorem, and there is a difference between contraction mapping and an inclusive mapping. Strictly speaking, one needs to distinguish functions from mappings and correspondences as well.

When the mapping is not weakly decreasing but only strictly decreasing, then the unique limit might be obtained in a finite time. If the mapping is strictly decreasing as in contraction mapping, then the sequence of images obtained by applying the contraction mapping to the initial set converges to the limit but never reaches the limit in finite time. However, as one can see from the description so far, and if anyone is familiar with the way these mappings work, one only iterates the same mapping to the image of the set until the image is sufficiently close to the limit. This is where the theoretical limit and numerical simulations does make a difference. Theoretically, there exists a unique limit that the sequence must converge to. Numerically, given that this limit is never reached in finite time, one needs to decide when the image is sufficiently close to the limit and one can take the outcome of the numerical simulation as the set of equilibrium payoffs or payoffs one can implement in a dynamic mechanism.

In addition to choosing bandwidths for error sizes and the stopping time, one should note that there is just one mapping that needs to be iterated. This is where the one-stage deviation principle in the usual sense of mechanism design matters. When the one-stage deviation principle in the usual sense holds in the given model, then there exists a corresponding (weakly) decreasing mapping one can iterate on. However, as I show in "Mechanism Design with Moral Hazard," (Kwon (2019) in short), the one-stage deviation principle in the usual sense of mechanism design doesn't hold in every model. In most papers, this step is shown individually. When the sufficient conditions I provide in Kwon (2019) hold, then one can apply the dynamic programming as usual. But when they don't apply, one needs to verify that the dynamic IC constraint is satisfied after every private history; the usual dynamic programming doesn't allow for this possibility, and the dynamic programming developed in this paper accomodates the dynamic IC constraint. When the one-stage deviation principle in the usual sense of mechanism design holds and when the usual dynamic programming holds, then the dynamic programming developed in this paper reaches the limit after the first iteration along the second dimension. To put it differently, if one were to index the iteration in my dynamic programming by a two-dimensional vector (i, T), then as the set converges to the limit with $i \to \infty$ at $T = 1$, the dynamic programming found the limit set, and one can stop the iteration without checking $T > 1$.

One should think of the iteration along T as verifying there is no profitable $T-$period deviation. This is why if the one-stage deviation principle in the usual sense holds, then there is no profitable deviation of longer lengths. The one-stage deviation principle in the sense of repeated games checks for deviations after every private history including deviations. In order to implement it in dynamic programming, it is implicit that the agent or players go back to the strategy the mechanism designer or other players are expecting. In macro, the representative agent might have private information. In any model without the informed-principal problem, the mechanism designer doesn't have any private information. The one-stage deviation principle in the usual sense means that the agent takes actions or makes reports as if he didn't deviate, i.e., as if his private information in the period he deviated is what the mechanism designer thinks was the agent's private information. In repeated games, most of the papers including APS focus on (imperfect) public monitoring. When players have private information, possibily due to private monitoring, APS doesn't generalize immediately without taking into account private histories and multi-period deviations.

As for examples of when one-stage deviation principle in the usual sense doesn't hold and one might need to worry about the dynamic IC, consider any stochastic game with private information of players. Any repeated game with private monitoring and a discount factor strictly bounded away from one also requires taking care of this issue. If the agent can have private information and also take an action privately, then dynamic mechanism design also needs to deal with multi-period deviations in general. The two sufficient conditions I provide in Kwon (2019) are (i) learning models with symmetric uncertainty and (ii) agents observing the payoff-relevant state perfectly every period in the Markovian environment. Otherwise, in any model with a fully-persistent state that is the agent's private information, one cannot rule out multi-period deviations a priori.

The rest of the paper is organized as follows. Section 2 describes the model, and section 3 presents results. Section 4 concludes.

2 Model

The current model starts with common prior for the mechanism designer and the agent, and only the agent's action is his private information. Results in section 3 are developed for this model, but one can expand the dynamic programming to allow for private signals, messages and recommendations. Given that this already leads to transfinite induction in dynamic programming, I will state all results for this model. But further private information just requires modifying the transfinite induction and the sequence of weakly decreasing mappings to be applied.

There exist one mechanism designer and one agent for $t = 1, 2, \cdots, \infty$. The common discount factor is $\delta \in (0, 1)$. Each period, there is a payoff-relevant state $\omega_t \in \Omega$, and neither the mechanism designer nor the agent receives a private signal. The common prior in the beginning of the first period is denoted by π^0, and in general, $\pi(\omega)$ is the belief on state ω. In period t, the agent takes an action $p_t \in A$ which is his private information, and an outcome $y_t \in Y$ is realized and observed by both parties. The mechanism designer makes a payment at the end of the period. I assume a Markovian environment with endogenous state which will be described formally in two paragraphs.

The conditions in this paragraph ensure existence of optimal mechanism; they are automatically satisfied in any finite environment. The set of states Ω, the set of actions A and the set of outcomes Y are non-empty compact Borel subsets of Polish (complete, separable, metric) spaces. $P_{\omega\omega'}(p)$, the probability of going from state ω to ω' when the agent chooses a, is a probablility measure on ω' given ω, p; P is jointly continuous in ω, ω', p. The cost of action $c(p) \in \mathbb{R}$ is continuous in p. $f_\omega(y)$ is the pdf of outcome y in state ω, i.e., measurable, non-negative and $\int_Y f_\omega(y)dy = 1$ for all $\omega \in \Omega$. Assume $f_\omega(y)$ is a continuous function of ω, y and $\int_Y y f_\omega(y)dy < M$ for some M, all $\omega \in \Omega$. Also assume full support, atomless, positive density everywhere. Lastly, assume f, P are such that if we start with a uniformly bounded common prior π in the beginning of first period, then resulting beliefs in all subsequent periods are also uniformly bounded. Or more precisely, sup norm is well-defined for resulting beliefs. If there are a finitely many states, we don't need to worry about it.

The distribution of outcome and the transition probabilities of the state are functions of ω, p this period. Denote the outcome distribution and the state transition at the end of period t by f_t, P_t, respectively:

$$f_t : \Omega \times A \to \sigma(Y),$$
$$P_t : \Omega \times A \to \sigma(\Omega)$$

where $\sigma(X)$ denotes the set of all probability measures on X. When $A \subseteq \mathbb{R}$, the cost of action for the agent is strictly increasing, strictly convex.

The mechanism designer values outcome y_t with $v : Y \to \mathbb{R}$. I assume the mechanism designer is risk neutral with respect to the payment w_t, and the agent values w_t with vNM utility function $u : \mathbb{R} \to \mathbb{R}$. u is strictly increasing, weakly concave. There is limited liability, but the lower bound need not be 0; this is due to a techanical reason and any bound that is sufficiently low works.

I will specify the agent's private history and the public history, which coincides with the mechanism designer's history, but I focus on IC constraints of the agent's action in this paper. With any detectable deviation, the public history between the mechanism designer and the agent contains a deviation as soon as the deviation occurs, and undetectable multi-period deviations are irrelevant. With limited commitment, the mechanism designer offers a mechanism at the beginning of each period, and the agent decides whether to accept or reject; given that this is observed by both the mechanism designer and the agent, I do not focus on any deviation at that stage.

If the agent doesn't participate or if the mechanism designer doesn't offer a mechanism, outside options for the mechanism designer and the agent are $\bar{v}$ and $\bar{u}$, respectively. With limited commitment, these are per-period, and with full commitment, these are the outside options in the first period. With limited commitment, the state transition in a period while they receive outside options is given by P^0, and the equilibrium notion is perfect Bayesian equilibrium.

Since with full commitment, the game is over when the agent doesn't participate, I define histories only for the case when the mechanism designer offers the mechanism on the equilibrium path, and the agent participates; the mechanism and the participation decision are omitted from histories. The mechanism consists of history-contingent payments.

The private history of the agent in period t is

$$h^{t,a} = (p_1, y_1, w_1, \cdots, p_{t-1}, y_{t-1}, w_{t-1}) \in \mathcal{H}^{t,a}.$$

The public history in the beginning of period t is $h^t = (y_1, w_1, \cdots, y_{t-1}, w_{t-1}) \in \mathcal{H}^t$. The strategy of the agent is $\sigma^{t,a} : \mathcal{H}^{t,a} \to A$. The allocation is $\sigma^t : \mathcal{H}^t \to \mathbb{R}$. The agent plays pure strategies, which is without loss of generality when $A \subseteq \mathbb{R}$ and the cost function is strictly convex, and the mechanism designer doesn't randomize over allocations. The mechanism designer is allowed to randomize over continuation contracts, and there is a public randomization device. All strategies are measurable functions. Throughout the paper, (h^t, h^k) denotes history h^t followed by h^k.

3 Results

Section 3.1 summarizes the relevant results for dynamic IC from Kwon (2019), and section 3.2 develops the dynamic programming method to find all numerical solutions. The code in C is in section B in the appendix. Section 3.3 presents simulations including how to pick stopping times, bandwidths for errors and so forth.

3.1 Quick Summary of Dynamic IC

Instead of deriving dynamic IC constraint and showing it to be necessary and sufficient condition for all IC constraints, I will summarize the relevant results from Kwon (2019) in this section.

The one-stage deviation IC constraints that have been commonly used in the dynamic mechanism design literature are different from what is referred to by the same name in the repeated games literature. In the dynamic mechanism design literature, it typically means deviating in one period or an instant then conforming to the mechanism designer's expectation from the following period or instant. When the agent reports his private information, this refers to lying only one period or instant then reporting truthfully from the following period or instant. If the agent takes an action which is unobserved by the mechanism designer, then it depends on the equilibrium strategy the mechanism designer expects. I show in Kwon (2019) that this types of one-stage deviation IC constraints are not always sufficient for all IC constraints. In particular, if the agent doesn't know the payoff-relevant state and only knows his past action, then it is in general not sufficient. When the agent privately observes the payoff-relevant state, it is still not sufficient as long as the agent's past actions or the past realizations of the payoff-relevant state matters for the continuation game.

When one-stage deviation ICs in the usual sense are no longer sufficient, then verifying incentive compatibility typically requires the knowledge of the agent's off-the-equilibrium-path strategy. I characterize an alternative way of verifying incentive compatibility only with the agent's equilibrium strategy in

Kwon (2019). This is the dynamic IC constraint, but it requires transfinite induction to implement in dynamic programming. What I mean by transfinite induction here is that the designer now needs to check for multi-period deviations of length k, instead of $k = 1$ as usual, and one can just check for all finite-length deviations, but this is practically not feasible.

I start with length 1 and characterize when the designer can stop checking for any longer chains. In numerical simulations, even if a sequence does converge in the limit, there is no guarantee that an element of the sequence will be the value of the limit at any finite index. Therefore, one needs to decide how close one wants the numerical value that comes out of dynamic programming to the theoretical limit, and with transfinite induction, there is an infinite sequence whose element is the limit of infinite sequence at each finite index. Therefore, one needs to pick the bandwidth or the size of error for each infinite sequence corresponding to an element in the main infinite sequence; then for the main sequence, it's similar to the dynamic programming that's been already studied.

As for the multi-period deviation IC constraints, the k-th element of the main infinite sequence checks for all deviation strategies of length k. Therefore, choosing the bandwidth for the k-th element of the main infinite sequence is the same as checking there is no k-period deviation that gives any profit beyond the chosen bandwidth. The transfinite induction starts with the first element of the main sequence, which coincides with the usual dynamic programming of deviating once and conforming to the designer's expectation afterwards. When there can be profitable multi-period deviations, now one needs to augment the usual dynamic programming by taking it as an element of infinite sequence; each element checks for k-period deviations. Therefore, choosing the bandwidth or the error size for the main infinite sequence is the same as there is no profitable deviation of any length that gives more profit than the chosen bandwidth. Practically, this bandwidth for the main infinite sequence translates into at which k, the designer can stop the second dimension of transfinite induction, i.e., there is no need to check for any deviation strategy of longer length.

I use a short-hand notation that drops the public history from the argument. For instance, $w(y)$ refers to the payment conditional on the public history up until that point and one more outcome realization y.

After outcome y is observed, the belief is updated from $\pi(\cdot)$ to

$$\pi^0(\omega) = \frac{\pi(\omega) f_\omega(y)}{\int_\Omega \pi(\omega') f_{\omega'}(y) d\omega'},$$

and in the following period, the belief is

$$\tilde{\pi}(\hat{\omega}) = \int_\Omega \frac{\pi(\omega) f_\omega(y) P_{\omega\hat{\omega}}(p)}{\int_\Omega \pi(\omega') f_{\omega'}(y) d\omega'} d\omega.$$

For π^0 to be well defined, $f_\omega(y)$ needs to be measurable in ω. For $\tilde{\pi}$ to be well defined, P given $\hat{\omega}, p$ needs to be measurable in ω.

Since P is jointly continuous in ω, p and Ω is compact, for given $\hat{\omega}$ and $\delta > 0$, there exists $\epsilon(\hat{\omega}, \delta) > 0$ such that $|P_{\omega\hat{\omega}}(p') - P_{\omega\hat{\omega}}(p)| < \delta$ for all $|p' - p| < \epsilon(\hat{\omega}, \delta)$. Since P is also continuous in $\hat{\omega}$, we can find $\epsilon(\hat{\omega}, \delta)$ continuous in $\hat{\omega}$, and together with the compactness of Ω, we get $\epsilon(\delta) > 0$ such that $|\tilde{\pi}_{p'}(\hat{\omega}) - \tilde{\pi}_p(\hat{\omega})| < \delta$ for all $\hat{\omega}$, $|p' - p| < \epsilon(\delta)$, and $\tilde{\pi}$ is a continuous function of p. $\tilde{\pi}$ is a continuous function of π^0. (P is jointly continuous on a compact set) When Ω, Y are compact and $f_\omega(y)$ is continuous in ω, y, π^0 is a continuous function of π. Up to here, I used pointwise convergence and sup norm.

The hypothetical continuation value of the agent is

$$\int_\Omega \int_Y -c(p) + u(w(y)) + \delta \int_\Omega V(y, \hat{\omega}) P_{\omega\hat{\omega}}(p) d\hat{\omega} f_\omega(y) dy \pi(\omega) d\omega.$$

In order for $u(w(y))$ to be measurable in y and $V(y, \omega)$ to be measurable in y, ω, it is enough that the mechanism designer offers $w(y)$ as a measurable function of y, and by our definition of $V(y, \omega)$, it should be measurable in both y, ω. Limited liability ensures $w(y) \geq -\underline{M}$ for some $\underline{M}$ sufficiently large.

The one-stage deviation IC in the usual sense of mechanism design is equivalent to

$$\begin{array}{rc} & \int_\Omega \int_Y -c(p) + u(w(y)) + \delta \int_\Omega V(y,\hat{\omega}) P_{\omega\hat{\omega}}(p) d\hat{\omega} f_\omega(y) dy \pi(\omega) d\omega \\ \geq & \int_\Omega \int_Y -c(p') + u(w(y)) + \delta \int_\Omega V(y,\hat{\omega}) P_{\omega\hat{\omega}}(p') d\hat{\omega} f_\omega(y) dy \pi(\omega) d\omega \\ \Leftrightarrow & \int_\Omega \int_Y c(p') - c(p) + \delta \int_\Omega V(y,\hat{\omega})(P_{\omega\hat{\omega}}(p) - P_{\omega\hat{\omega}}(p')) d\hat{\omega} f_\omega(y) dy \pi(\omega) d\omega \geq 0 \end{array}$$

If we assume $P_{\omega\hat{\omega}}(p)$ is differentiable in p, we can take the left and right limits (after dividing by $p' - p$) and get the equality constraint.

The dynamic IC is given by

$$\sum_{n=t}^{\infty} \delta^{n-t} \int \int_\Omega \int_Y c(p'_n(\tilde{h^n})) - c(p_n(\hat{h^n})) + \delta \int_\Omega V_{y\hat{\omega}} (P_{\omega\hat{\omega}}(p_n(\hat{h^n})) - P_{\omega\hat{\omega}}(p'_n(\tilde{h^n}))) d\hat{\omega} f_\omega(y) dy \tilde{\pi^n}(\omega) d\omega dG \geq$$

where G is the cdf of reaching each history given the agent's true private history.

3.2 Dynamic Programming

When the on-path single deviation IC is sufficient, the standard dynamic programming can allow for adverse selection or ex-ante symmetric uncertainty together with moral hazard. However, when the dynamic IC is necessary, i.e., one must account for multi-period deviations, the standard dynamic programming no longer works. In order to characterize the optimal mechanism only with the dynamic IC, one needs to make sure that the dynamic IC is also sufficient. This happens when there is limited commitment or continuity at infinity. I will describe the intuition for dynamic programming when the dynamic IC is necessary and sufficient in this section. The formal proofs are in the appendix A.

The standard dynamic programming starts with the candidate set of payoffs (typically the set of all individually rational payoffs) and apply a contraction mapping until it reaches the fixed point. Existing literature has focused on cases when on-path single deviation ICs are sufficient, and when they are sufficient, the contraction mapping corresponds to the on-path single deviation given the public history up to that point.

With the dynamic IC and multi-period deviations, it is no longer sufficient to apply one contraction mapping until it reaches the fixed point. There is a sequence of operators, which are monotone but not necessarily contraction mappings, and the dynamic programming requires trans-finite induction on this sequence of operators. However, since each operator leads to a monotone sequence of set of payoffs, one can still start with the candidate set of payoffs and apply each operator until it reaches the fixed point. The difference from the standard dynamic programming is that once it reaches the fixed point with the N-th operator T_N, it goes on to apply the $(N+1)$-th operator T_{N+1} until it reaches the next fixed point. Figure 1 shows how the trans-finite induction works, where $W^{N,0}$ is the initial set for T_N and $W^{N,k} = (T_N)^k(W^{N,0})$.

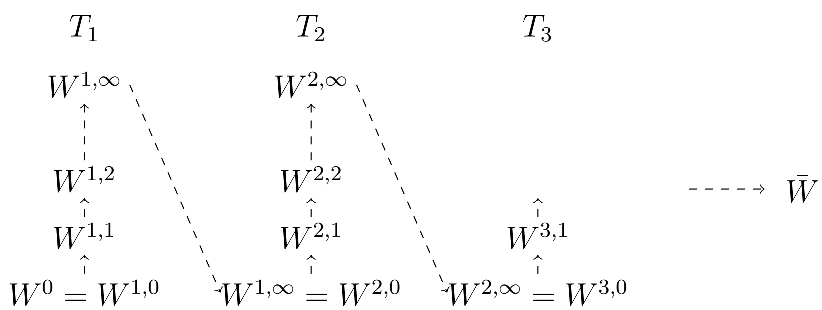

Figure 1: Trans-finite Induction

The state variable for dynamic programming is $(\pi, V(\cdot,\cdot), V^P)$ which takes into account the belief on the state, hypothetical continuation values for the agent and the supremeum of principal's payoffs, all on the equilibrium path. With one agent, the dynamic IC only requires hypothetical continuation values on the equilibrium path, and there is only one dynamic IC after every history on the equilibrium path; it is sufficient to take the variables on the equilibrium path. Multi-period deviations are taken care of within each dynamic IC, and we don't need to worry about belief disagreement off the equilibrium path for the state variable.

In the first stage, the candidate set of payoffs is $W^0 = \{(\pi, V(\cdot,\cdot), V^P) || V(y,\omega)| \leq \bar{V}\}$ for some $\bar{V}$. All that matters at this point is that there is a uniform upper bound on hypothetical continuation values of the agent, since otherwise, no matter how many times one applies monotone mappings to the initial set, it might never converge to a set with finite measure. With limited commitment or continuity at infinity, the agent's expected utility on the equilibrium path is uniformly bounded. But we still need to show that hypothetical continuation values are bounded; hypothetical continuation values integrate up to a finite expected utility, but in principle, they can still diverge or even be infinite on a set of measure zero. The proof follows from the fact that the hypothetical continuation value is continuous in ω and a, and any continuous function on a compact set is uniformly bounded.

Let T_1 be the operator with the on-path single deviation IC. One can find all points in W that can be generated by W, and this corresponds to the usual mapping in the standard dynamic programming except that it is weakly decreasing and not strictly a contraction mapping. Since $T_1(W) \subseteq W$, $(T_1)^j(W^0) = W^{1,j}$ converges to a set-theoretic limit. We can define T_N to be the mapping with the N-period IC and use the fact that the dynamic IC is equivalent to satisfying the N-period IC for every N. Each T_N is non-increasing, and the limit of a monotone sequence is well-defined.

For each T_N, N-period deviations are taken care of as follows. First, given π^t, we know π^{t+1} when the agent takes action a and outcome y is realized. By induction, we can choose $V_{t+2}(y, \omega)$ for each π^{t+1} such that from period $t+1$ on, $(N-1)$-period deviation ICs are satisfied. Next, choose $w_{t+1}(y)$ such that $V_{t+1}(y, \omega)$ generated by $w_{t+1}(y)$, $V_{t+2}(y, \omega)$ satisfy IR and the promise-keeping constraint. We know the equilibrium beliefs, actions for the next N periods, and we need to verify the N-period deviation IC.

N-period deviation IC is satisfied by backward induction. Let $\tilde{V}_{t+k+1}$ be the agent's maximum deviation payoff from any deviation between period $t+k+1$ and $t+N-1$; the agent conforms to the principal's expectation from period $t+N$ onwards. In period $t+N-1$, given the agent's private belief $\tilde{\pi}^{t+N-1}$ and his hypothetical continuation values $V_{t+N}(y, \omega)$, the agent has the optimal action. In period $t+k$, given the agent's private belief $\tilde{\pi}^{t+k}$ and the maximum deviation payoff from $t+k+1$ on, $\tilde{V}_{t+k+1}$, the agent has the optimal action. We can continue doing the backward induction, and in period t, we have $\tilde{\pi}^t = \pi^t$ and the equilibrium payoff has to be weakly better than the most profitable deviation payoff. The rest of the argument follows from the agent's deviation payoff being a continuous function of his private belief and action; the set of beliefs and the set of actions are compact sets.

Since I take care of all private beliefs the agent might have in period $t + k$ when I verify the N-period deviation IC, there is no need to keep track of the agent's private belief as the state variable.

We want the expected utility of the agent to be a bounded upper semi-continuous function of his belief and his action. Hypothetical continuation values $V(y, \omega)$ can be thought of as a bounded function $V : Y \times \Omega \to \mathbb{R}$. It is bounded because the principal has no commitment power and the expected outcome in each state is uniformly bounded. If the principal has full-commitment power, I need to show that the expected utility is bounded. (this is necessary both for the agent's optimal action to be well-defined and also for continuity at infinity) I'll use the fact that the dynamic IC (which implies the local IC) is equivalent to satisfying any N-period IC for all N.

Lemma 1. *Hypothetical continuation values on the equilibrium path are uniformly bounded.*

Theorem 1. *The dynamic programming is well-defined, i.e., there exists a sequence of set operations such that the set-theoretic limit is the largest self-generating set. The agent's optimal actions for any N-period IC in the largest self-generating set is well-defined. The supremum of the principal's payoff is well-defined.*

3.3 Simulations

This section will be included in the submitted version.

4 Conclusion

I develop a dynamic programming method to implement the dynamic IC constraint from Kwon (2019). Kwon (2019) characterizes a necessary and sufficient condition for all IC constraints when the one-stage deviation principle in the usual sense of mechanism design doesn't hold in a given model. The dynamic IC becomes necessary in many environments when the agent doesn't observe the payoff-relevant state every period and can take an action privately.

Most of dynamic programming methods implemented in the literature works under the one-stage deviation principle in the usual sense of mechanism design. When the agent deviates, he only deviates once and goes back to the strategy the mechanism designer "expects" given the public history. In repeated-games literature, issues I point out in Kwon (2019) are irrelevant as long as the monitoring technology is (imperfect) public monitoring.

Once the underlying environment one wants to study doesn't satisfy the one-stage deviation principle in the usual sense of mechanism design, then there is a need for dynamic programming to accomodate undetectable multi-period deviations. The method I develop in this paper involves transfinite induction, and as long as one can verify there is no profitable and undetectable multi-period deviations, the dynamic programming method need not be unique.

References

[1] Abreu, Dilip, and Yuliy Sannikov. "An algorithm for twoplayer repeated games with perfect monitoring." *Theoretical Economics* 9.2 (2014): 313-338.

[2] Abreu, Dilip, Benjamin Brooks, and Yuliy Sannikov. "A "Pencil-Sharpening" Algorithm for Two Player Stochastic Games with Perfect Monitoring." *Princeton University William S. Dietrich II Economic Theory Center Research Paper* 078_2016 (2016).

[3] Abreu, Dilip, David Pearce, and Ennio Stacchetti. "Optimal cartel equilibria with imperfect monitoring." *Journal of Economic Theory* 39.1 (1986): 251-269.

[4] Abreu, Dilip, David Pearce, and Ennio Stacchetti. "Toward a theory of discounted repeated games with imperfect monitoring." *Econometrica* 58.5 (1990): 1041-1063.

A Proofs

Proof of Lemma 1. We know that the equilibrium payoffs are uniformly bounded, but a priori, we cannot rule out having unbounded hypothetical continuation values on a set of measure zero; I restrict attention to equivalent classes that coincide on a set of measure one. I'll show that $V(y_{t-1}, \omega_t)$ is a continuous function of ω_t and p_t. Given $V(y_{t-1}, \omega_t)$, we have $w(y_t), V_{t+1}(y_t, \omega_{t+1})$ such that $V(y_{t-1}, \omega_t) = \int_Y -c(p_t) + u(w(y_t)) + \delta \int_\Omega V(y_t, \omega_{t+1}) P_{\omega_t \omega_{t+1}}(p_t) d\omega_{t+1} f_{\omega_t}(y_t) dy_t$. I'll first show that $\int_\Omega V(y_t, \omega_{t+1}) P_{\omega_t \omega_{t+1}}(p_t) d\omega_{t+1}$ is continuous in ω_t, p_t. Define $g(\omega, p) = \int_\Omega V(y, \omega') P_{\omega\omega'}(p) d\omega'$ and consider a sequence $\omega^n \to \omega$. Since P is jointly continuous, $V(y, \omega') P_{\omega^n \omega'}(p) \to V(y, \omega') P_{\omega\omega'}(p)$ almost everywhere. Together with the compactness of Ω, the joint continuity of P implies that for given ω, there exists a neighborhood $\mathcal{N}_\omega$ such that $P_{\omega^n \omega'}(p) \leq \kappa P_{\omega\omega'}(p)$ for some $\kappa > 0$ and all ω', p and $\omega^n \in \mathcal{N}_\omega$. We already know that $V(y, \omega') P_{\omega\omega'}(p)$ is integrable, and limited liability implies that hypothetical continuation values are bounded from below. Then there exists a neighborhood $\mathcal{N}'_\omega \subseteq \mathcal{N}_\omega$ such that $|V(y, \omega')|$ is also integrable on $\mathcal{N}'_\omega$. By the dominated convergence theorem, $g(\omega^n, p) \to g(\omega, p)$. The proof for continuity in p_t is similar. The proof of continuity of $V(y_{t-1}, \omega_t)$ in ω_t, p_t is similar, and we use the fact that $-c(p_t) + u(w(y_t)) + \delta \int_\Omega V(y_t, \omega_{t+1}) P_{\omega_t \omega_{t+1}}(p_t) d\omega_{t+1} f_{\omega_t}(y_t)$ is bounded from below. Therefore, the hypothetical continuation value is a continuous function on a compact set and is bounded. □

Proof of Theorem 1. I will set up the dynamic programming problem and show that at each iteration, the most profitable deviation for the agent is well-defined. The sequence of sets we get after each iteration is non-increasing and has a well-defined limit in the set-theoretic sense. The largest self-generating set is non-empty because the agent choosing the cheapest action and the principal making no payment is an equilibrium. With the compact action set and the continuous cost function, the cheapest action exists. With no commitment, the relevant constraints for the dynamic programming are (i) the principal offers the equilibrium contract (ii) the agent accepts/rejects according to the equilibrium strategy (iii) the agent's dynamic IC (iv) the principal makes the payment (v) the promise-keeping constraint. With within-period commitment, the relevant constraints for the dynamic programming are (i) the principal offers the equilibrium contract (ii) the agent accepts/rejects according to the equilibrium strategy (iii) the agent's dynamic IC (iv) the promise-keeping constraint. Whether the principal has no commitment power or within-period commitment power matters for the minmax NE. I'm not going to specify the minmax NE here, but if either the principal or the agent prefers his outside option over the minmax NE, then they'll take their outside options. This pins down the lower bound on payoffs for the IR constraints. For the rest of the proof, I assume within-period commitment power and ignore the principal's incentives to make payments he promised; with no commitment power, this will put an upper bound on the payment the principal can make (the continuation payoff minus the minmax NE or the outside option). Computationally, I can just impose the outside options and see the minmax NE from the largest self-generating set. If the minmax NE is better than the outside option for both the principal and the agent, then I need to use the minmax NE instead of the outside options. I could also just assume that the minmax NE we get by imposing IR with outside options is worse than taking their outside options.

I will construct a sequence of operations so that the limit is the largest self-generating set we want. I don't think I need the agent's optimal action to be unique, but I still need the agent to play a pure strategy. The state space for the dynamic programming is $(\pi, V(\cdot,\cdot), V^P$ where V^P is the principal's expected payoff. The argument I'm going to use for non-local ICs should work as long as there is continuity at infinity. If there is continuity at infinity, there must be a profitable N-period deviation for N sufficiently large, and we can do backward induction.

I need to specify the sequence of operations: I start with the local IC, and for each N, I iterate the operation for the N-period deviation IC until I reach the limit. Once I have the limit for N-period deviations, I continue with $(N+1)$-period deviations. And I take the limit as $N \to \infty$. Let's start with $W^0$where $V(\cdot,\cdot)$ are just assumed to be bounded by the uniform bound on the hypothetical continuation values. The iteration for the N-period deviation IC is T_N, and the limit of T_N starting with W^{N-1} is W^N. Also define $T_N(W^{N,i-1}) = W^{N,i}$, $W^{N,0} = W^{N-1}$. T_1 is just the standard largest self-generating set with the local IC constraint. Among the constraints, (i) and (ii) just mean that the payoffs are weakly greater than the outside options (or min-max NE). (iv) can be taken care of as follows: Suppose we have V_{t+2}. When we choose $w_{t+1}(y)$ for each y, we can pin down V_{t+1} that is consistent with the promise-keeping constraint. First find the largest self-generating set subject to (ii), (iv) and the local IC (without worrying about the principal's payoff). Once we have W^1, we know V^P for each pair of $(\pi, V(\cdot,\cdot))$ and can keep only those that satisfy (i). If there are multiple V^Ps corresponding to $(\pi, V(\cdot,\cdot))$ then choose the supremum of V^P (following the principle of optimality). At this point, we haven't shown that whether the supremum can be obtained as the maximum. But we also know that once we have V^P we can generate any $\hat{V^P} < V^P$ as long as it's weakly greater than the principal's outside option. For T_1, we can show that the agent's optimal action is well-defined because the agent's expected utility is bounded and is a continuous function of his action. Generally speaking, to show that the most profitable N-period deviation is well-defined, I need to show a version of selection theorem, and I need Ω to be a Borel subset of a Polish space, A to be a compact metric space and the agent's expected utility from an N-period deviation to be bounded and upper semi-continuous.

T_N for $N \geq 2$ are defined as follows: Fix π^t and an action p, and we can find the beliefs π^{t+1} that are consistent with π^t, p. Choose $V_{t+2}(y,\omega)$ from $W^{N,i}$ and $w_{t+1}(y)$ for each π^{t+1} such that V_{t+1}, V^P_{t+1} given by the promise-keeping constraint also satisfies (i), (ii) and the local IC constraint for period t is satisfied at p. By construction, there are no profitable $(N-1)$-period deviations starting with $V_{t+2}(y,\omega)$, and we can find optimal actions for the next $N-1$ periods and $V_{t+N}(y,\omega)$ from period $t+N$ on. We should have the payments after each history from period $t+1$ to period $t+N-1$. We also know the equilibrium belief after each history. Fix $\pi^{t+\tilde{N}-1}$ for each history and we can find the most profitable deviation for the agent in $t+N-1$ and therefore assign the agent's maximum deviation payoff from period $t+N-1$

on as a function of $\pi^{\widetilde{t+N-1}}$ given the continuation game from $t+N-1$ on the equilibrium path. Fix $\pi^{\widetilde{t+N-2}}$ then conditional on the agent's action p'_{t+N-2} we know the agent's beliefs $\pi^{\widetilde{t+N-1}}$and his maximum deviation payoff. We can find the most profitable deviation for the agent in period $t+N-2$ and assign the agent's deviation payoff from $t+N-2$ on as a function of $\pi^{\widetilde{t+N-2}}$ and the continuation game from $t+N-2$ on the equilibrium path. We can repeat this until we reach $\pi^{\tilde{t}}$. To show that the most profitable deviation is well-defined, suppose we are in period $t+n$ with $\pi^{\widetilde{t+n}}, p_{t+n}, w_{t+n}(\cdot), \tilde{V}_{t+n+1}(\cdot,\cdot)$ where $\tilde{V}_{t+n+1}$ is the agent's maximum deviation payoff from period $t+n+1$ on. (in period $t+N-1$, these will just be the hypothetical continuation values from period $t+N$ on) In period $t+n$, the agent's expected utility from period $t+n$ on is a continuous function of $\pi^{\widetilde{t+n}}, p_{t+n}$ and we know that it is finite for $\pi^{\widetilde{t+n}} = \pi^{t+n}$ and the equilibrium action p_{t+n}. The proof follows the proof of Lemma 1 closely, and we know from limited liability that $\tilde{V}_{t+n+1}$ is bounded from below. Since there is no profitable $(N-1)$-period deviation, if the agent starts with π^{t+n} and chooses p_{t+n}, his maximum deviation payoffs from the next period on coincides with his equilibrium payoffs; it follows that the agent's expected utility from period $t+n$ on is bounded. Since the set of priors and the set of actions is compact, we know that the product of the two is compact (Tychonoff's Theorem). Therefore, the agent's deviation payoff from period $t+n$ on is a continuous function on a compact set and is bounded. Therefore, for given $\pi^{\widetilde{t+n}}$, there exists the maximum deviation payoff for the agent, and the agent has the most profitable deviation. But in period t, the principal and the agent share the same prior $\pi^{\tilde{t}} = \pi^t$. Keep $V_{t+1}(y,\omega)$ for π^t if and only if it is incentive compatible with respect to the maximum deviation payoff. (I use backward induction to find the most profitable N-period deviation for the agent, but I also use backward induction to show that the maximum deviation payoff for the agent is a continuous function of his belief and his action in the given period. This needs a proof because the agent's maximum deviation payoff from the next period on depends on his belief in the next period)

Since each operation $T_N : W^{N,i-1} \to W^{N,i}$ satisfies $W^{N,i} = T(W^{N,i-1}) \subseteq W^{N,i-1}$, we have a monotone sequence, and the limit is well-defined in the set-theoretic sense. By construction, it is the largest self-generating set satisfying all four conditions. It also follows from the previous paragraph that the agent's optimal actions for any N-period IC is well-defined. The supremum of the principal's payoff for any given π is well-defined. □

B Dynamic Programming Code in C

This will be included in the submitted version.