Supplement to "Dynamic contracting: an irrelevance theorem"

Péter Eső^{*} and Balázs Szentes[†]

December 21, 2015

Online Appendix

Proof of Proposition 2

Since in a pure adverse selection model $\tilde{u}_{tat} \equiv 0$ for all t, throughout this section, we remove a_t from the arguments of \tilde{u}_t , that is, $\tilde{u}_t : \Theta_t \times X^t \to \mathbb{R}$. We first inspect the consequences of Assumptions 1 and 2 on the orthogonalized model. Note that since θ_t does not depend on x_{t-1} , the ψ_t inference functions do not depend on the decisions either, so $\psi_t : \mathcal{E}^t \to \Theta_t$. The time-separability of the agent's payoff (part (i) of Assumption 2) is preserved in the orthogonalized model, except that the flow utility at $t, u_t : \mathcal{E}^t \times X^t \to \mathbb{R}$, now depends on the history of types up to and including time t:

$$u_t\left(\varepsilon^t, x^t\right) = \widetilde{u}_t\left(\psi_t\left(\varepsilon^t\right), x^t\right). \tag{1}$$

Part (iii) of Assumption 1 implies that the larger the type history in the orthogonalized model up to time t, the larger is the corresponding period-t type in the original model. This, coupled with part (ii) of Assumption 2 implies that u_t is weakly increasing in ε^{t-1} and strictly in ε_t . Monotonicy in x^t as well as single-crossing (part (iii) of Assumption 2) are also preserved in the orthogonalized model. We state these properties formally in the following Lemma.

Lemma 1. (i) For all $t \in \{0, ..., T\}$ and $\hat{\varepsilon}^t, \varepsilon^t \in \mathcal{E}^t$,

$$\widehat{\varepsilon}^{t} \leq \varepsilon^{t} \Rightarrow \psi_{t}\left(\widehat{\varepsilon}^{t}\right) \leq \psi_{t}\left(\varepsilon^{t}\right), \qquad (2)$$

and the inequality is strict whenever $\hat{\varepsilon}_t < \varepsilon_t$.

 $^{\ ^*} Department of Economics, Oxford University, Oxford, UK. \ E-mail: \ peter.eso@economics.ox.ac.uk.$

[†]Department of Economics, London School of Economics, London, UK. E-mail: b.szentes@lse.ac.uk.

(ii) The flow utility, u_t defined by (1), is weakly increasing in ε^{t-1} and x^{t-1} , and strictly increasing in x_t and ε_t .

(iii) For all $t \in \{0, ..., T\}$, $u_{t\varepsilon_t}(\varepsilon^t, x^t) \ge u_{t\varepsilon_t}(\varepsilon^t, \widehat{x}^t)$ whenever $x^t \ge \widehat{x}^t$.

Proof. Part (i) $\varepsilon_t = H_t^{-1}(\theta_t | \theta_{t-1})$, therefore $\psi_0(\varepsilon_0) = H_0^{-1}(\varepsilon_0)$ and ψ_t for t > 0 is defined recursively by $\psi_t(\varepsilon^t) = H_t^{-1}(\varepsilon^t | \psi^{t-1}(\varepsilon^{t-1}))$. We prove the statement of this part by induction. For t = 0, we have $H_0^{-1}(\varepsilon_0) \ge H_0^{-1}(\widehat{\varepsilon}_0)$ whenever $\varepsilon_0 \ge \widehat{\varepsilon}_0$ and the inequality is strict if $\varepsilon_0 > \widehat{\varepsilon}_0$.

Suppose that (2) holds for t, that is, $\psi_t \left(\widehat{\varepsilon}^t \right) \leq \psi_t \left(\varepsilon^t \right)$ whenever $\widehat{\varepsilon}^t \leq \varepsilon^t$ and the inequality is strict whenever $\widehat{\varepsilon}^t < \varepsilon_t$. Note that $\psi_{t+1} \left(\widehat{\varepsilon}_{t+1} \right) = H_{t+1}^{-1} \left(\widehat{\varepsilon}_{t+1} | \psi_t \left(\widehat{\varepsilon}^t \right) \right)$ and $\psi_{t+1} \left(\varepsilon^{t+1} \right) =$ $H_{t+1}^{-1} \left(\varepsilon_{t+1} | \psi_t \left(\varepsilon^t \right) \right)$. Since $\psi_t \left(\widehat{\varepsilon}^t \right) \leq \psi_t \left(\varepsilon^t \right)$ by the inductive hypothesis, part (ii) of Assumption 1 implies that $\psi_{t+1} \left(\widehat{\varepsilon}^{t+1} \right) \leq \psi_{t+1} \left(\varepsilon^{t+1} \right)$. In addition, if $\varepsilon_{t+1} > \widehat{\varepsilon}_{t+1}$ then $H_{t+1}^{-1} \left(\varepsilon_{t+1} | \psi_t \left(\varepsilon^t \right) \right) >$ $H_{t+1}^{-1} \left(\varepsilon'_{t+1} | \psi_t \left(\widehat{\varepsilon}^t \right) \right)$.

Part (ii): The function u_t is strictly increasing in x_t and weakly increasing x^{t-1} because of part (ii) Assumption 2 and (1). Equalities (1) and (2) imply that u_t is strictly increasing in ε_t and weakly increasing in ε^{t-1} .

Part (iii): Fix a $t \in \{0, ..., T\}$ and note that by (1),

$$u_{t\varepsilon_{t}}\left(\varepsilon^{t}, x^{t}\right) = \widetilde{u}_{t\theta_{t}}\left(\psi_{t}\left(\varepsilon^{t}\right), x^{t}\right) \frac{\partial\psi_{t}\left(\varepsilon^{t}\right)}{\partial\varepsilon_{t}}.$$

The result follows from (2) and part (iii) of Assumption 2. \Box

Another important consequence of part (i) of Assumption 1 is that for all ε^{t+1} and $\widehat{\varepsilon}_t$, there exists a type $\sigma_{t+1}(\varepsilon^{t+1}, \widehat{\varepsilon}_t) \in \mathcal{E}_t$ such that, fixing the principal's past and future decisions as well as the realizations of the agent's types beyond period t + 1, the agent's utility flow from period t + 1 on is the same with type history ε^{t+1} as it is with $(\varepsilon^{t-1}, \widehat{\varepsilon}_t, \sigma_{t+1}(\varepsilon^{t+1}, \widehat{\varepsilon}_t))$. We will show below that σ_{t+1} , interpreted in Eső and Szentes (2007) as the agent's "correction of a lie," defines an optimal strategy for the agent at time t+1 after a deviation from truthtelling in an incentive compatible direct mechanism at t. This is formally stated in the following

Lemma 2. For all $t \in \{0, \ldots, T-1\}$, $\varepsilon^{t+1} \in \mathcal{E}^{t+1}$ and $\widehat{\varepsilon}_t \in \mathcal{E}_t$, there exists a unique $\sigma_{t+1}(\varepsilon^{t+1}, \widehat{\varepsilon}_t) \in \mathcal{E}_{t+1}$ such that for all $k = t+1, \ldots, T$, all $\widehat{\varepsilon}^k \in \mathcal{E}^k$ and $\widehat{x}^k \in X^k$,

$$u_k(\varepsilon^{t-1},\varepsilon_t,\varepsilon_{t+1},\widehat{\varepsilon}_{t+2},\ldots,\widehat{\varepsilon}_k,\widehat{x}^k) = u_k(\varepsilon^{t-1},\widehat{\varepsilon}_t,\sigma_{t+1},\widehat{\varepsilon}_{t+2},\ldots,\widehat{\varepsilon}_k,\widehat{x}^k).$$
(3)

The function σ_{t+1} is increasing in ε_t , strictly increasing in ε_{t+1} and decreasing in $\hat{\varepsilon}_t$. **Proof.** Fix a $t \in \{0, \ldots, T-1\}, \varepsilon^{t+1} \in \mathcal{E}^{t+1}$ and $\hat{\varepsilon}_t \in \mathcal{E}_t$. Let

$$\sigma_{t+1} = H_{t+1} \left(\psi_{t+1} \left(\varepsilon^{t+1} \right) | \psi_t \left(\varepsilon^{t-1}, \widehat{\varepsilon}_t \right) \right).$$
(4)

By the full support assumption in part (i) of Assumption 1, it follows that

$$\psi_{t+1}\left(\varepsilon^{t+1}\right) = \psi_{t+1}\left(\varepsilon^{t-1}, \widehat{\varepsilon}_t, \sigma_{t+1}\right),$$

that is, the computed time-(t+1) type of the original model is the same after ε^{t+1} and $(\varepsilon^{t-1}, \widehat{\varepsilon}_t, \sigma_{t+1})$. Therefore the inferred type in the original model is also the same after any future observations, that is,

$$\psi_k\left(\varepsilon^{t-1},\varepsilon_t,\varepsilon_{t+1},\widehat{\varepsilon}_{t+2},\ldots,\widehat{\varepsilon}_k\right)=\psi_k(\varepsilon^{t-1},\widehat{\varepsilon}_t,\sigma_{t+1},\widehat{\varepsilon}_{t+2},\ldots,\widehat{\varepsilon}_k),$$

for all k = t + 1, ..., T, all $\hat{\varepsilon}^k \in \mathcal{E}^k$. This equality and (1) imply (3). Also note that $\sigma_{t+1}(\varepsilon^{t+1}, \hat{\varepsilon}_t)$, defined by (4), is increasing in ε_t , strictly increasing in ε_{t+1} by part (i) of Lemma 1 and decreasing in $\hat{\varepsilon}_t$ by part (i) of Lemma 1 and part (iii) of Assumption 1.

It remains to show that there does not exist any other σ_{t+1} which satisfies (3). This follows from part (ii) of Lemma 1, which states that u_{t+1} is strictly increasing in ε_{t+1} , which implies that (3) with k = t + 1 cannot hold for two different σ_{t+1} 's. \Box

The statement of the previous lemma might appear somewhat complicated at first glance, but its meaning and its intuitive proof are quite straightforward. Part (i) of Assumption 1 requires the support of θ_t to be independent of θ_{t-1} . Therefore, if the type of the agent is $\psi_t (\varepsilon^{t-1}, \widehat{\varepsilon}_t)$ at time t, there is a chance that the period-(t+1) type will be $\psi^{t+1} (\varepsilon^{t+1})$. The type $\sigma_{t+1} (\varepsilon^{t+1}, \widehat{\varepsilon}_t)$ denotes the orthogonalized information of the agent at t+1 which induces the transition from $\psi_t (\varepsilon^{t-1}, \widehat{\varepsilon}_t)$ to $\psi_{t+1} (\varepsilon^{t+1})$, that is,

$$\psi_{t+1}\left(\varepsilon^{t-1},\widehat{\varepsilon}_{t},\sigma_{t+1}\left(\varepsilon^{t+1},\widehat{\varepsilon}_{t}\right)\right)=\psi_{t+1}\left(\varepsilon^{t+1}\right).$$

This means that the inferred type in the original model is the same after the histories $(\varepsilon^{t-1}, \widehat{\varepsilon}_t, \sigma_{t+1} (\varepsilon^{t+1}, \widehat{\varepsilon}_t))$ and ε^{t+1} . Part (i) of Assumption 1 and part (ii) of Assumption 2 imply that, given the decisions, the flow utilities in the future only depend on current type which, in turn, imply (3).

The decision rule in the orthogonalized model, $\{x_t : \mathcal{E}^t \to X_t\}_{t=0}^T$, which corresponds to $\{\tilde{x}_t\}_{t=0}^T$, is defined by $x_t(\varepsilon^t) = \tilde{x}_t(\psi^t(\varepsilon^t))$ for all t and ε^t . Note that, by (2), if $\{\tilde{x}_t\}_{t=0}^T$ is increasing in type (\tilde{x}_t is increasing in θ^t for all t) then the corresponding decision rule $\{x_t\}_{t=0}^T$ in the orthogonalized model is also increasing in type.¹

In fact, the monotonicity of $\{\widetilde{x}_t\}_{t=0}^T$ implies a stronger monotonicity condition on $\{x_t\}_{t=0}^T$. Consider the following two type histories, ε^k and $(\varepsilon_1, ..., \varepsilon_{t-1}, \widehat{\varepsilon}_t, \sigma_{t+1}(\varepsilon^{t+1}, \widehat{\varepsilon}_t), \varepsilon_{t+2}, ..., \varepsilon_k)$.

¹To see this, note that if $v^t \ge \hat{v}^t$ then $x_t \left(\hat{v}^t \right) = \tilde{x}_t \left(\psi^t \left(\hat{v}^t \right) \right) \le \tilde{x}_t \left(\psi^t \left(v^t \right) \right) = x_t \left(v^t \right)$, where the inequality follows from the monotonicity of $\{ \tilde{x}_t \}_0^T$ and (2).

Note that the inferred types in the original model are exactly the same along these histories except at time t. At time t, the inferred type is smaller after ε^t if and only if $\varepsilon_t \leq \hat{\varepsilon}_t$. Since \tilde{x}_k is increasing in θ_t , the decision is smaller after ε^k if and only if $\varepsilon_t \leq \hat{\varepsilon}_t$. Formally,

Remark 1. If $\{\widetilde{x}_t\}_{t=0}^T$ is increasing then for all k = 1, ..., T, t < k, $\varepsilon^k \in \mathcal{E}^k$:

$$x^{k}(\varepsilon^{k}) \leq x^{k}\left(\varepsilon^{t-1}, \widehat{\varepsilon}_{t}, \sigma_{t+1}\left(\varepsilon^{t+1}, \widehat{\varepsilon}_{t}\right), \varepsilon_{t+2}, ..., \varepsilon_{k}\right) \Leftrightarrow \widehat{\varepsilon}_{t} \geq \varepsilon_{t}.$$
(5)

Proof of Remark 1. Recall from the proof of Lemma 2 that for all k = t + 1, ..., T,

$$\psi_k\left(\varepsilon^{t+1},\varepsilon_{t+2}...,\varepsilon_k\right) = \psi_k\left(\varepsilon^{t-1},\widehat{\varepsilon}_t,\sigma_{t+1}\left(\varepsilon^{t+1},\widehat{\varepsilon}_t\right),\varepsilon_{t+2},...,\varepsilon_k\right)$$

By (2), $\psi_t(\varepsilon^t) \leq \psi_t(\varepsilon^{t-1}, \widehat{\varepsilon}_t)$ if and only if $\widehat{\varepsilon}_t \geq \varepsilon_t$. Then (5) follows from the monotinicity of $\{\widetilde{x}_t\}_0^T$ and the definition of $\{x\}_0^T$. \Box

To simplify the exposition, we introduce the following notation for $t = 0, \ldots, T, k \ge t$:

$$\zeta_t^k(\varepsilon^k, y) = \left(\varepsilon^{t-1}, y, \varepsilon_{t+1}, ..., \varepsilon_k\right),$$

$$\rho_t^k(\varepsilon^k, y, \widehat{\varepsilon}_t) = \left(\varepsilon^{t-1}, \widehat{\varepsilon}_t, \sigma_{t+1}\left(\varepsilon^{t-1}, y, \varepsilon_{t+1}, \widehat{\varepsilon}_t\right), \varepsilon_{t+2}, ..., \varepsilon_k\right).$$

The vectors $\zeta_t^k(\varepsilon^k, y)$ and $\rho_t^k(\varepsilon^k, y, \widehat{\varepsilon}_t)$ are type histories up to period k, true or reported, which are different from ε^k only at t or at t and t + 1. For k = t these are appropriately truncated, e.g., $\rho_t^t(\varepsilon^t, y, \widehat{\varepsilon}_t) = (\varepsilon^{t-1}, \widehat{\varepsilon}_t)$.

As we explained, the monotonicity of $\{\tilde{x}_t\}_{t=0}^T$ implies both the monotonicity of $\{x_t\}_{t=0}^T$ and (5). Therefore, in order to prove Proposition 2, it is sufficient to show that any increasing decision rule in the orthogonalized model which satisfies (5) can be implemented. In what follows, fix a direct mechanism with an increasing decision rule $\{x_t\}_{t=0}^T$ that satisfies (5). Let $\Pi_t(\varepsilon_t|\varepsilon^{t-1})$ denote a truthful agent's expected payoff at t conditional on ε^t . That is,

$$\Pi_t(\varepsilon_t|\varepsilon^{t-1}) = E\left[\sum_{k=0}^T u_k\left(\varepsilon^k, x^k(\varepsilon^k)\right) - p(\varepsilon^T) \middle| \varepsilon^t\right].$$
(6)

Define the payment function, p, such that for all t = 0, ..., T and $\varepsilon^t \in \mathcal{E}^t$,

$$\Pi_t \left(\varepsilon_t | \varepsilon^{t-1} \right) = \Pi_t \left(0 | \varepsilon^{t-1} \right) + E \left[\int_0^{\varepsilon_t} \sum_{k=t}^T u_{k\varepsilon_t} \left(\zeta_t^k (\varepsilon^k, y), x^k \left(\zeta_t^k (\varepsilon^k, y) \right) \right) dy \bigg| \varepsilon^t \right].$$
(7)

It is not hard to show that the integral on the right-hand side of (7) exists and is finite because of part (ii) of Assumption 1, part (i) of Assumption 2 and the monotonicity of x^k . It should be clear that it is possible to define p such that (7) holds. In this mechanism, let $\pi_t (\varepsilon_t, \widehat{\varepsilon}_t | \varepsilon^{t-1})$ denote the expected payoff of the agent at time t whose type history is ε^t and has reported $(\varepsilon^{t-1}, \widehat{\varepsilon}_t)$. This is the maximum payoff she can achieve from using any reporting strategy from t + 1 conditional on the type history ε^t and on the reports $(\varepsilon^{t-1}, \widehat{\varepsilon}_t)$. If the mechanism is incentive compatible then, clearly, $\Pi_t(\varepsilon_t | \varepsilon^{t-1}) = \pi_t(\varepsilon_t, \varepsilon_t | \varepsilon^{t-1})$.

We call a mechanism *IC after time* t if, conditional on telling the truth before and at time t - 1, it is an equilibrium strategy for the agent to tell the truth afterwards, that is, from period t on. By Lemma 2, the continuation utilities of the agent with type ε^{t+1} are the same as those of the agent with type $(\varepsilon^{t-1}, \hat{\varepsilon}_t, \sigma_{t+1}(\varepsilon^{t+1}, \hat{\varepsilon}_t))$ conditional on the reports and the realization of types after t + 1. Therefore, if a mechanism is IC after t + 1, the agent whose type history is ε^{t+1} and reported $(\varepsilon^{t-1}, \hat{\varepsilon}_t)$ up to time t maximizes her continuation payoff by reporting $\sigma_{t+1}(\varepsilon^{t+1}, \hat{\varepsilon}_t)$ at time t + 1 and reporting truthfully afterwards. If this were not the case then the agent with $(\varepsilon^{t-1}, \hat{\varepsilon}_t, \sigma_{t+1}(\varepsilon^{t+1}, \hat{\varepsilon}_t))$ would have a profitable deviation after truthful reports up to and including t, contradicting the assumption that the mechanism is IC after t + 1. Therefore, in a mechanism that is IC after t + 1, we have

$$\pi_t \left(\varepsilon_t, \widehat{\varepsilon}_t | \varepsilon^{t-1} \right) = u_t \left(\varepsilon^t, x^t \left(\varepsilon^{t-1}, \widehat{\varepsilon}_t \right) \right) - u_t \left(\varepsilon^{t-1}, \widehat{\varepsilon}_t, x^t \left(\varepsilon^{t-1}, \widehat{\varepsilon}_t \right) \right) + \int \Pi_{t+1} \left(\sigma_{t+1} \left(\varepsilon^{t+1}, \widehat{\varepsilon}_t \right) | \varepsilon^{t-1}, \widehat{\varepsilon}_t \right) d\varepsilon_{t+1}.$$
(8)

We use (8) in the following Lemma to characterize the agent's continuation payoff who deviates at t in a mechanism that is IC after t.

Lemma 3. Suppose that the mechanism is IC after time t + 1 and (7) is satisfied. Then, for all ε^t and $\widehat{\varepsilon}_t$,

$$\pi_t\left(\varepsilon_t,\widehat{\varepsilon}_t|\varepsilon^{t-1}\right) - \pi_t\left(\widehat{\varepsilon}_t,\widehat{\varepsilon}_t|\varepsilon^{t-1}\right) = \sum_{k=t}^T E\left[\int_{\widehat{\varepsilon}_t}^{\varepsilon_t} u_{k\varepsilon_t}\left(\zeta_t^k(\varepsilon^k,y), x^k\left(\rho_t^k(\varepsilon^k,y,\widehat{\varepsilon}_t)\right)\right)dy\bigg|\varepsilon^t\right].$$
 (9)

This lemma is a direct generalization of Lemma 5 of Eső and Szentes (2007).

Proof. Let $\gamma_t^k \left(\varepsilon^k, \widehat{\varepsilon}_t, y \right)$ denote $\left(\varepsilon^{t-1}, \widehat{\varepsilon}_t, y, \varepsilon_{t+2}, ..., \varepsilon_k \right)$ for k = t+1, ..., T. Suppose first that

 $\varepsilon_t > \widehat{\varepsilon}_t$. Then $\sigma_{t+1} \left(\varepsilon^{t+1}, \widehat{\varepsilon}_t \right) > \varepsilon_{t+1}$, and

$$\begin{aligned} \pi_t \left(\varepsilon_t, \widehat{\varepsilon}_t | \varepsilon^{t-1} \right) &= u_t \left(\varepsilon^t, x^t \left(\varepsilon^{t-1}, \widehat{\varepsilon}_t \right) \right) - u_t \left(\varepsilon^{t-1}, \widehat{\varepsilon}_t, x^t \left(\varepsilon^{t-1}, \widehat{\varepsilon}_t \right) \right) \\ &+ \int \Pi_{t+1} \left(\sigma_{t+1} \left(\varepsilon^{t+1}, \widehat{\varepsilon}_t \right) | \varepsilon^{t-1}, \widehat{\varepsilon}_t \right) d\varepsilon_{t+1} \\ &= u_t \left(\varepsilon^t, x^t \left(\varepsilon^{t-1}, \widehat{\varepsilon}_t \right) \right) - u_t \left(\varepsilon^{t-1}, \widehat{\varepsilon}_t, x^t \left(\varepsilon^{t-1}, \widehat{\varepsilon}_t \right) \right) + \Pi_t \left(\widehat{\varepsilon}_t | \varepsilon^{t-1} \right) \\ &+ \sum_{k=t+1}^T \int \dots \int \int_{\varepsilon_{t+1}}^{\sigma_{t+1} \left(\varepsilon^{t+1}, \widehat{\varepsilon}_t \right)} u_{\varepsilon_{t+1}}^k \left(\gamma_t^k \left(\varepsilon^k, \widehat{\varepsilon}_t, y \right), x^k \left(\gamma_t^k \left(\varepsilon^k, \widehat{\varepsilon}_t, y \right) \right) \right) dy d\varepsilon_{t+1} \dots d\varepsilon_k \\ &= u_t \left(\varepsilon^t, x^t \left(\varepsilon^{t-1}, \widehat{\varepsilon}_t \right) \right) - u_t \left(\varepsilon^{t-1}, \widehat{\varepsilon}_t, x^t \left(\varepsilon^{t-1}, \widehat{\varepsilon}_t \right) \right) + \pi_t \left(\widehat{\varepsilon}_t, \widehat{\varepsilon}_t | \varepsilon^{t-1} \right) \\ &+ \sum_{k=t+1}^T \int \dots \int \int_{\varepsilon_{t+1}}^{\sigma_{t+1} \left(\varepsilon^{t+1}, \widehat{\varepsilon}_t \right)} u_{k\varepsilon_{t+1}} \left(\gamma_t^k \left(\varepsilon^k, \widehat{\varepsilon}_t, y \right), x^k \left(\gamma_t^k \left(\varepsilon^k, \widehat{\varepsilon}_t, y \right) \right) \right) dy d\varepsilon_{t+1} \dots d\varepsilon_k \end{aligned}$$

where the first equality is just (8), the second one follows from (7), and the third one from $\Pi_t(\widehat{\varepsilon}_t|\varepsilon^{t-1}) = \pi_t(\widehat{\varepsilon}_t,\widehat{\varepsilon}_t|\varepsilon^{t-1})$. So, in order to prove (9), we only need to show that

$$u_t\left(\varepsilon^t, x^t\left(\varepsilon^{t-1}, \widehat{\varepsilon}_t\right)\right) - u_t\left(\varepsilon^{t-1}, \widehat{\varepsilon}_t, x^t\left(\varepsilon^{t-1}, \widehat{\varepsilon}_t\right)\right) = \int_{\widehat{\varepsilon}_t}^{\varepsilon_t} u_{t\varepsilon_t}\left(\varepsilon^t_{-t}, y, x^t\left(\rho_t^t(\varepsilon^t, \widehat{\varepsilon}_t)\right)\right) dy.$$
(10)

and

$$\sum_{k=t+1}^{T} \int \dots \int \int_{\varepsilon_{t+1}}^{\sigma_{t+1}\left(\varepsilon^{t+1},\widehat{\varepsilon}_{t}\right)} u_{k\varepsilon_{t+1}}\left(\gamma_{t}^{k}\left(\varepsilon^{k},\widehat{\varepsilon}_{t},y\right), x^{k}\left(\gamma_{t}^{k}\left(\varepsilon^{k},\widehat{\varepsilon}_{t},y\right)\right)\right) dyd\varepsilon_{t+1}...d\varepsilon_{k}$$
$$= \sum_{k=t+1}^{T} \int \dots \int \int_{\widehat{\varepsilon}_{t}}^{\varepsilon_{t}} u_{t\varepsilon_{t}}\left(\zeta_{t}^{k}(\varepsilon^{k},y), x^{k}\left(\rho_{t}^{k}(\varepsilon^{k},y,\widehat{\varepsilon}_{t})\right)\right) dyd\varepsilon_{t+1}...d\varepsilon_{k}.$$
(11)

Equation (10) directly follows from the Fundamental Theorem of Calculus. We now turn our attention to (11). By Lemma 2, σ_{t+1} is continuous and monotone. The image of $\sigma_{t+1} (\varepsilon^{t+1}, y)$ on $y \in [\widehat{\varepsilon}_t, \varepsilon_t]$ is $[\varepsilon_{t+1}, \sigma_{t+1} (\varepsilon^{t+1}, \widehat{\varepsilon}_t)]$. Hence, after changing the variables of integration, for all k = t + 1, ..., T:

$$\int_{\varepsilon_{t+1}}^{\sigma_{t+1}(\varepsilon^{t+1},\widehat{\varepsilon}_{t})} u_{k\varepsilon_{t+1}} \left(\gamma_{t}^{k} \left(\varepsilon^{k}, \widehat{\varepsilon}_{t}, y \right), x^{k} \left(\gamma_{t}^{k} \left(\varepsilon^{k}, \widehat{\varepsilon}_{t}, y \right) \right) \right) dy = \int_{\widehat{\varepsilon}_{t}}^{\varepsilon_{t}} u_{k\varepsilon_{t+1}} \left(\gamma_{t}^{k} \left(\varepsilon^{k}, \widehat{\varepsilon}_{t}, \sigma_{t+1} \left(\varepsilon^{t-1}, y, \varepsilon_{t+1}, \widehat{\varepsilon}_{t} \right) \right), x^{k} \left(\gamma_{t}^{k} \left(\varepsilon^{k}, \widehat{\varepsilon}_{t}, \sigma_{t+1} \left(\varepsilon^{t-1}, y, \varepsilon_{t+1}, \widehat{\varepsilon}_{t} \right) \right) \right) \right) \\ \frac{\partial \sigma_{t+1} \left(\varepsilon^{t-1}, y, \varepsilon_{t+1}, \widehat{\varepsilon}_{t} \right)}{\partial y} dy. \quad (12)$$

Recall that by (3) the following is an identity in y:

$$u_k\left(\varepsilon^{t-1}, y, \varepsilon_{t+1}, ..., \varepsilon_k, x^k\right) \equiv u_k\left(\gamma_t^k\left(\varepsilon^k, \widehat{\varepsilon}_t, \sigma_{t+1}\left(\varepsilon^{t-1}, y, \varepsilon_{t+1}, \widehat{\varepsilon}_t\right)\right), x^k\right),$$

so, by the Implicit Function Theorem,

$$u_{k\varepsilon_{t}}\left(\varepsilon^{t-1}, y, \varepsilon_{t+1}, ..., \varepsilon_{k}, x^{k}\right) = u_{k\varepsilon_{t+1}}\left(\varepsilon^{t-1}, \widehat{\varepsilon}_{t}, \sigma_{t+1}\left(\varepsilon^{t-1}, y, \varepsilon_{t+1}, \widehat{\varepsilon}_{t}\right), ..., \varepsilon_{k}, x^{k}\right) \frac{\sigma_{t+1}\left(\varepsilon^{t-1}, y, \varepsilon_{t+1}, \widehat{\varepsilon}_{t}\right)}{\partial y}.$$
 (13)

Plugging (13) into (12) and noting that $\gamma_t^k \left(\varepsilon^k, \widehat{\varepsilon}_t, \sigma_{t+1} \left(\varepsilon^{t-1}, y, \varepsilon_{t+1}, \widehat{\varepsilon}_t \right) \right) = \rho_t^k \left(\varepsilon^k, y, \widehat{\varepsilon}_t \right)$ yields (11).

An identical argument can be used to deal with the case where $\hat{\varepsilon}_t > \varepsilon_t$.

We are now ready to prove Proposition 2.

Proof of Proposition 2. In order to prove that the transfers defined by (7) implement $\{x_t\}_{t=0}^T$, it is enough to prove that the mechanism is IC after all t = 0, ..., T - 1. We prove this by induction. For t = T - 1 this follows from the standard result in static mechanism design with the observation that x_T is monotone and (7) is satisfied for T. Suppose now that the mechanism is IC after t + 1. We show that the mechanism is IC after t, that is, the agent has no incentive to lie at t if she has told the truth before t.

Consider an agent with type history ε^t and report history ε^{t-1} who is contemplating to report $\hat{\varepsilon}_t < \varepsilon_t$. We have to show that $\pi_t (\varepsilon_t, \hat{\varepsilon}_t | \varepsilon^{t-1}) - \pi_t (\varepsilon_t, \varepsilon_t | \varepsilon^{t-1}) \leq 0$ which can be written as

$$\pi_t\left(\varepsilon_t,\widehat{\varepsilon}_t|\varepsilon^{t-1}\right) - \pi_t\left(\widehat{\varepsilon}_t,\widehat{\varepsilon}_t|\varepsilon^{t-1}\right) + \pi_t\left(\widehat{\varepsilon}_t,\widehat{\varepsilon}_t|\varepsilon^{t-1}\right) - \pi_t\left(\varepsilon_t,\varepsilon_t|\varepsilon^{t-1}\right) \le 0.$$

By (7) and (9), the previous inequality can be expressed as

$$\sum_{k=t}^{T} E\left[\int_{\widehat{\varepsilon}_{t}}^{\varepsilon_{t}} u_{k\varepsilon_{t}}\left(\zeta_{t}^{k}(\varepsilon^{k}, y), x^{k}\left(\zeta_{t}^{k}(\varepsilon^{k}, y)\right)\right) dy \middle| \varepsilon^{t}\right]$$

$$\geq \sum_{k=t}^{T} E\left[\int_{\widehat{\varepsilon}_{t}}^{\varepsilon_{t}} u_{k\varepsilon_{t}}\left(\zeta_{t}^{k}(\varepsilon^{k}, y), x^{k}\left(\rho_{t}^{k}(\varepsilon^{k}, y, \widehat{\varepsilon}_{t})\right)\right) dy \middle| \varepsilon^{t}\right].$$

$$(14)$$

In order to prove this inequality it is enough to show that the integrand on the left-hand side is larger than the integrand on the right-hand side. By part (iii) of Lemma 1, in order to show this, we only need that $x^k \left(\rho_t^k\left(\varepsilon^k, y, \widehat{\varepsilon}_t\right)\right) \leq x^k \left(\zeta_t^k(\varepsilon^k, y)\right)$ on $y \in [\widehat{\varepsilon}_t, \varepsilon_t]$, which follows from Remark 1. An identical argument can be used to rule out deviation to $\widehat{\varepsilon}_t > \varepsilon_t$. \Box

From the proof of Proposition 2 it is clear that in the environment satisfying Assumptions

1 and 2 (i.e., with Markov types and a well-behaved agent payoff function), a decision rule $\{\tilde{x}_t\}_{t=0}^T$ is implemented by transfers satisfying (7) if, and only if, condition (14) holds in the orthogonalized model.² But (7) is also a necessary condition of implementation (differentiate it in ε_t and compare that with the condition in Proposition 1), therefore condition (14) is indeed the necessary and sufficient condition of implementability of a decision rule in the regular, Markov environment. Formally, we state

Remark 2. Suppose that Assumptions 1 and 2 hold. Then a decision rule, $\{\tilde{x}_t\}_0^T$, is implementable if, and only if, (14) holds in the model with orthogonalized information.

Implementability in the Benchmark Case. — Suppose that the principal can observe $\varepsilon_1, ..., \varepsilon_T$. Then, using arguments in standard static mechanism design, a decision rule $\{x_t\}_0^T$ can be implemented if, and only if, for all $\widehat{\varepsilon}_0, \varepsilon_0 \in \mathcal{E}_0, \widehat{\varepsilon}_0 \leq \varepsilon_0$,

$$E\left[\sum_{k=0}^{T}\int_{\widehat{\varepsilon}_{0}}^{\varepsilon_{0}}u_{k\varepsilon_{0}}\left(y,\varepsilon_{-0}^{k},x^{k}\left(y,\varepsilon_{-0}^{k}\right)\right)dy\bigg|\varepsilon_{0}\right] \geq E\left[\sum_{k=0}^{T}\int_{\widehat{\varepsilon}_{0}}^{\varepsilon_{0}}u_{k\varepsilon_{0}}\left(y,\varepsilon_{-0}^{k},x^{k}\left(\widehat{\varepsilon}_{0},\varepsilon_{-0}^{k}\right)\right)dy\bigg|\varepsilon_{0}\right]$$

This inequality is obviously a weaker condition than (14), so the principal can implement more allocations in the benchmark case.

Proof of Proposition 5

In order to be able to refer to the additional restrictions required by the proposition, we state the strict single-crossing properties in the following

Assumption 6.

(i) For all $t \in \{0, ..., T\}$, $\theta_t \in \Theta_t$, $a_t \in A_t$: $\widetilde{u}_{t\theta_t} (\theta_t, a_t, x^t) > \widetilde{u}_{t\theta_t} (\theta_t, a_t, \widehat{x}^t)$ whenever $x^t > \widehat{x}^t$.

(ii) For all $t \in \{0, ..., T\}$, for all $\theta_t \in \Theta_t, a_t \in A_t, x^t \in X^t \ \widetilde{u}_{t\theta_t x_\tau} \left(\theta_t, a_t, x^t\right) f_{ta_t} \left(\theta_t, a_t\right) > \widetilde{u}_{ta_t x_\tau} \left(\theta_t, a_t, x^t\right) f_{t\theta_t} \left(\theta_t, a_t\right).$

Recall that in the proof of Proposition 4 we decomposed the gain from any deviation strategy into the sum of two parts. The first part is the difference between the payoff from deviating and truth-telling in the hypothetical model where y is contractible. The second part is the difference between the payoff from the misreporting strategy but matching the actions with the misreports and the payoff from misreporting and altering the actions optimally. Then we appealed to Proposition 2 to conclude that the first part is negative and proved that the second part is small. The key to the proof of this proposition is to show that if

²Note that condition (14) is a joint restriction on $\{x_t\}_0^T$ and the marginal utility of the agent's type, and it is implied by the monotonicity of the decision rule in the environment of Assumptions 1-2.

the deviation strategy leads to a decision rule which is far away from $(\tilde{\mathbf{x}}^T, \tilde{\mathbf{y}}^T)$, then the first part of the decomposed payoff is not only negative but also large relative to the possible gain corresponding to the second part. In order to do so, we follow the standard argument in static mechanism design to estimate the deviation payoffs. This estimation is based on the single-crossing property and we have derived the key formula, (14), in the orthogonalized model.

In what follows, we use the notation introduced in the proof of Proposition 4. Recall that the payoff difference corresponding to the first part of the decomposition is

$$E_{\theta^{T}}\left[\sum_{t=0}^{T} w_{t}\left(\theta_{t}, \widetilde{\mathbf{y}}_{t}\left(\theta^{t}\right), \widetilde{\mathbf{x}}^{t}\left(\theta^{t}\right)\right) - \overline{\mathbf{p}}\left(\theta^{T}\right)|\theta_{0}\right]$$

$$-E_{\theta^{T}}\left[\sum_{t=0}^{T} w_{t}\left(\theta_{t}, \widetilde{\mathbf{y}}_{t}\left(\rho_{t}\left(\theta^{t}\right)\right), \widetilde{\mathbf{x}}^{t}\left(\rho_{t}(\theta^{t})\right) - \overline{\mathbf{p}}\left(\rho^{T}\left(\theta^{T}\right)\right)|\theta_{0}\right].$$

$$(15)$$

Since (14) is derived in the orthogonalized model, we rewrite the previous inequality in terms of the orthogonalized information structure. To this end, let $(\mathbf{\tilde{x}}_t, \mathbf{\tilde{a}}_t)_{t=0}^T$ denote the allocation corresponding to $(\mathbf{\tilde{x}}_t, \mathbf{\tilde{a}}_t)_{t=0}^T$, that is, $(\mathbf{\tilde{x}}_t (\varepsilon^t), \mathbf{\tilde{a}}_t (\varepsilon^t)) \equiv (\mathbf{\tilde{x}}_t (\psi^t (\varepsilon^t)), \mathbf{\tilde{a}}_t (\psi^t (\varepsilon^t)))$ for all t, ε^t . Similarly, define $\mathbf{\tilde{y}}_t (\varepsilon^t)$ and $\mathbf{\bar{p}}_t (\varepsilon^t)$ to be $\mathbf{\tilde{y}}_t (\psi^t (\varepsilon^t))$ and $\mathbf{\bar{p}} (\psi^t (\varepsilon^t))$, respectively, for all t and ε^t . Let $\underline{\rho}_t (\varepsilon^t)$ denote the deviation strategy, that is, $\underline{\rho}_t (\varepsilon^t) = \rho_t (\psi^t (\varepsilon^t))$. Finally, let $\omega_t (\varepsilon^t, y, x) \equiv w_t (\psi^t (\varepsilon^t), y, x)$ for all t = 0, ..., T. It is not hard to prove that by Assumption 6 it follows that there exists an $\widetilde{m} \in \mathbb{R}_+$, such that for all $t \in \{0, ..., T\}$, $\varepsilon_t \in \Theta_t$:

$$\omega_{t\varepsilon_t}\left(\varepsilon^t, y_t, x^t\right) - \omega_{t\varepsilon_t}\left(\varepsilon^t, \widehat{y}_t, \widehat{x}^t\right) \ge \widetilde{m} \left\| \left(y_t, x^t \right) - \left(\widehat{y}_t, \widehat{x}^t \right) \right\|$$
(16)

whenever $(y_t, x^t) \ge (\widehat{y}_t, \widehat{x}^t)$ and $(y_t, x^t), (\widehat{y}_t, \widehat{x}^t) \in \left\{ \left(\underline{\widetilde{y}}_t(\varepsilon^t), \underline{\widetilde{x}}^t(\varepsilon^t) \right) : \varepsilon^t \in \mathcal{E}^t \right\}$. Using these notations, we can rewrite (15) as

$$E_{\varepsilon^{T}}\left[\sum_{t=0}^{T}\omega_{t}\left(\varepsilon_{t},\underline{\widetilde{\mathbf{y}}}_{t}\left(\varepsilon^{t}\right),\underline{\widetilde{\mathbf{x}}}^{t}\left(\varepsilon^{t}\right)\right)-\underline{\overline{\mathbf{p}}}\left(\varepsilon^{T}\right)|\varepsilon_{0}\right]\\-E_{\varepsilon^{T}}\left[\sum_{t=0}^{T}\omega_{t}\left(\varepsilon_{t},\underline{\widetilde{\mathbf{y}}}_{t}\left(\rho_{t}\left(\varepsilon^{t}\right)\right),\underline{\widetilde{\mathbf{x}}}^{t}\left(\rho_{t}(\varepsilon^{t})\right)-\underline{\overline{\mathbf{p}}}\left(\rho^{T}\left(\varepsilon^{T}\right)\right)|\varepsilon_{0}\right].$$

Observe that it is without the loss of generality to assume that there exists a K > 0 such that

$$\left\| \left(\underline{\widetilde{\mathbf{y}}}_{t} \left(\varepsilon^{t} \right), \underline{\widetilde{\mathbf{x}}}^{t} \left(\varepsilon^{t} \right) \right) - \left(\underline{\widetilde{\mathbf{y}}}_{t} \left(\varepsilon^{\prime t} \right), \underline{\widetilde{\mathbf{x}}}^{t} \left(\varepsilon^{\prime t} \right) \right) \right\| \leq K \left\| \varepsilon^{t} - \varepsilon^{\prime t} \right\|$$
(17)

for all $t \in \{0, ..., T\}$ and $\varepsilon^t \in \mathcal{E}^t$ because any increasing allocation rule can be approximated

arbitrarily well in the L_2 -norm with a decision rule which satisfies the previous inequality. Also note that the set $\left\{ \left(\underline{\widetilde{\mathbf{y}}}_t \left(\varepsilon^t \right), \underline{\widetilde{\mathbf{x}}}^t \left(\varepsilon^t \right) \right) : \varepsilon^t \in \mathcal{E}^t \right\}$ is bounded. For notational convenience, we assume that

$$\left\| \left(\underline{\widetilde{\mathbf{y}}}_{t} \left(\varepsilon^{t} \right), \underline{\widetilde{\mathbf{x}}}^{t} \left(\varepsilon^{t} \right) \right) \right\| \leq \frac{1}{2}$$
(18)

for all $t \in \{0, ..., T\}$ and $\varepsilon^t \in \mathcal{E}^t$.

Next, we show that for all $\tilde{\delta} > 0$ and $\tau = 0, ..., T$ we can construct payment rules so that

$$E_{\varepsilon^{t}} \left\| \left(\underline{\widetilde{\mathbf{y}}}_{t} \left(\varepsilon^{t} \right), \underline{\widetilde{\mathbf{x}}}_{t} \left(\varepsilon^{t} \right) \right) - \left(\underline{\widetilde{\mathbf{y}}}_{t} \left(\rho \left(\varepsilon^{t} \right) \right) \right), \underline{\widetilde{\mathbf{x}}}_{t} \left(\rho \left(\varepsilon^{t} \right) \right) \right\|^{2} < \widetilde{\delta} / \left(T + 1 \right),$$

$$(19)$$

for all $t \leq \tau$. We prove this statement by induction. Consider $\tau = 0$. As we mentioned before, we use (14) to estimate the loss from a deviation. In particular, we estimate this loss by the difference between the first term of the summation in the left-hand side and the first term of the summation on the right-hand side of (14). In other words, we approximate the the loss due to a time t deviation by the instantaneous loss and ignore future losses. Hence, by (14),

$$\begin{split} E_{\varepsilon^{T}} \left[\sum_{t=0}^{T} \omega_{t} \left(\varepsilon_{t}, \underline{\widetilde{\mathbf{y}}}_{t} \left(\varepsilon^{t} \right), \underline{\widetilde{\mathbf{x}}}^{t} \left(\varepsilon^{t} \right) \right) - \underline{\overline{\mathbf{p}}} \left(\varepsilon^{T} \right) |\varepsilon_{0} \right] \\ - E_{\varepsilon^{T}} \left[\sum_{t=0}^{T} \omega_{t} \left(\varepsilon_{t}, \underline{\widetilde{\mathbf{y}}}_{t} \left(\underline{\rho}_{t} \left(\varepsilon^{t} \right) \right), \underline{\widetilde{\mathbf{x}}}^{t} \left(\underline{\rho}_{t} \left(\varepsilon^{t} \right) \right) - \underline{\overline{\mathbf{p}}} \left(\underline{\rho}^{T} \left(\varepsilon^{T} \right) \right) |\varepsilon_{0} \right] \right] \\ \geq E_{\varepsilon_{0}} \left[\int_{\varepsilon_{0}}^{\rho(\varepsilon_{0})} \omega_{0\varepsilon_{0}} \left(z, \underline{\widetilde{\mathbf{y}}}_{0} \left(z \right), \underline{\widetilde{\mathbf{x}}}^{0} \left(z \right) \right) - \omega_{0\varepsilon_{0}} \left(z, \underline{\widetilde{\mathbf{y}}}_{0} \left(\varepsilon_{0} \right) \right), \underline{\widetilde{\mathbf{x}}}^{0} \left(\underline{\rho}_{t} \left(\varepsilon_{0} \right) \right) \right) dz \right]. \end{split}$$

By (16), the right-hand side of the previous inequality is weakly larger than

$$\widetilde{m}E_{\varepsilon_{0}}\left[\int_{\varepsilon_{0}}^{\rho(\varepsilon_{0})}\left\|\left(\underline{\widetilde{\mathbf{y}}}_{0}\left(\varepsilon_{0}\right),\underline{\widetilde{\mathbf{x}}}^{0}\left(\varepsilon_{0}\right)\right)-\left(\underline{\widetilde{\mathbf{y}}}_{0}\left(\underline{\rho}_{t}\left(\varepsilon_{0}\right)\right),\underline{\widetilde{\mathbf{x}}}^{0}\left(\underline{\rho}_{t}\left(\varepsilon_{0}\right)\right)\right)\right\|\,dz\right].$$

Furthermore, by (17), the previous expression is larger that

$$\widetilde{m}E_{\varepsilon_{0}}\left[\frac{\left\|\left(\underline{\widetilde{\mathbf{y}}}_{0}\left(\varepsilon_{0}\right),\underline{\widetilde{\mathbf{x}}}^{0}\left(\varepsilon_{0}\right)\right)-\left(\underline{\widetilde{\mathbf{y}}}_{0}\left(\underline{\rho}_{t}\left(\varepsilon_{0}\right)\right),\underline{\widetilde{\mathbf{x}}}^{0}\left(\underline{\rho}_{t}\left(\varepsilon_{0}\right)\right)\right)\right\|^{2}}{2K}\right].$$

Recall that in the proof of Proposition 4, we proved that for each $\delta > 0$ it is possible to construct payments so that the second part of the decomposed gain from deviation is less

than δ . Therefore, in order to guarantee that the deviation $(\rho_t)_t$ is profitable it must be that

$$\widetilde{m}E_{\varepsilon_{0}}\left[\frac{\left\|\left(\underline{\widetilde{\mathbf{y}}}_{0}\left(\varepsilon_{0}\right),\underline{\widetilde{\mathbf{x}}}^{0}\left(\varepsilon_{0}\right)\right)-\left(\underline{\widetilde{\mathbf{y}}}_{0}\left(\underline{\rho}_{t}\left(\varepsilon_{0}\right)\right),\underline{\widetilde{\mathbf{x}}}^{0}\left(\underline{\rho}_{t}\left(\varepsilon_{0}\right)\right)\right)\right\|^{2}}{2K}\right]<\delta,$$

that is,

$$E_{\varepsilon_{0}}\left[\left\|\left(\underline{\widetilde{\mathbf{y}}}_{0}\left(\varepsilon_{0}\right),\underline{\widetilde{\mathbf{x}}}^{0}\left(\varepsilon_{0}\right)\right)-\left(\underline{\widetilde{\mathbf{y}}}_{0}\left(\underline{\rho}_{t}\left(\varepsilon_{0}\right)\right),\underline{\widetilde{\mathbf{x}}}^{0}\left(\underline{\rho}_{t}\left(\varepsilon_{0}\right)\right)\right)\right\|^{2}\right]<\frac{2K\delta}{\widetilde{m}}.$$
(20)

So, choosing $\delta = \widetilde{m}\widetilde{\delta}/\left(2K\left(T+1\right)\right)$ yields the claim in (19) for $\tau = 0$.

Suppose that the claim in (19) is true for $\tau \geq 0$. We show that this claim is also true for $\tau + 1$. The difficulty with the inductive step is that (14) can only be used to estimate the time- $(\tau + 1)$ loss due to a deviation if there were no deviations in previous periods. Let us explain how we overcome this problem. By the inductive hypothesis, there are payments so that the optimal deviation strategy induces a decision rule which is arbitrarily close to the decision rule generated by truth-telling in time periods $0, 1, ... \tau$. Therefore, we can approximate the optimal deviation by a misreporting strategy which specifies truth-telling until period τ and then coincides with the optimal deviation. According to this approximating deviation strategy, the first deviation occurs in period $\tau + 1$ and hence, we can use (14) to estimate the loss due to this deviation once again. To this end, let $\underline{\rho}_t^{\tau}(\varepsilon^t)$ be defined as follows

$$\underline{\rho}_{t}^{\tau}\left(\varepsilon^{t}\right) = \begin{cases} \varepsilon^{t} & \text{if } t \leq \tau\\ \underline{\rho}_{t}\left(\varepsilon^{t}\right) & \text{if } t > \tau, \end{cases}$$

that is, $\left\{\underline{\rho}_{t}^{\tau}\right\}_{t}$ is a deviation strategy which prescribes truth-telling until period t and, after

period t, it coincides with $\left\{\underline{\rho}_t\right\}_t$. Using this notation,

$$\begin{split} E_{\varepsilon^{T}} \left[\sum_{t=0}^{T} \omega_{t} \left(\varepsilon_{t}, \underline{\widetilde{\mathbf{y}}}_{t} \left(\varepsilon^{t} \right), \underline{\widetilde{\mathbf{x}}}^{t} \left(\varepsilon^{t} \right) \right) - \underline{\mathbf{p}} \left(\varepsilon^{T} \right) |\varepsilon_{0} \right] \\ - E_{\varepsilon^{T}} \left[\sum_{t=0}^{T} \omega_{t} \left(\varepsilon_{t}, \underline{\widetilde{\mathbf{y}}}_{t} \left(\underline{\rho}_{t} \left(\varepsilon^{t} \right) \right), \underline{\widetilde{\mathbf{x}}}^{t} \left(\underline{\rho}_{t} \left(\varepsilon^{t} \right) \right) - \underline{\mathbf{p}} \left(\underline{\rho}^{T} \left(\varepsilon^{T} \right) \right) |\varepsilon_{0} \right] \right] \\ = E_{\varepsilon^{T}} \left[\sum_{t=0}^{T} \omega_{t} \left(\varepsilon_{t}, \underline{\widetilde{\mathbf{y}}}_{t} \left(\varepsilon^{t} \right), \underline{\widetilde{\mathbf{x}}}^{t} \left(\varepsilon^{t} \right) \right) - \underline{\mathbf{p}} \left(\varepsilon^{T} \right) |\varepsilon_{0} \right] \\ - E_{\varepsilon^{T}} \left[\sum_{t=0}^{T} \omega_{t} \left(\varepsilon_{t}, \underline{\widetilde{\mathbf{y}}}_{t} \left(\underline{\rho}_{t}^{T} \left(\varepsilon^{t} \right) \right), \underline{\widetilde{\mathbf{x}}}^{t} \left(\underline{\rho}_{t}^{T} \left(\varepsilon^{t} \right) \right) - \underline{\mathbf{p}} \left(\underline{\rho}^{\tau T} \left(\varepsilon^{T} \right) \right) |\varepsilon_{0} \right] \\ + E_{\varepsilon^{T}} \left[\sum_{t=0}^{T} \omega_{t} \left(\varepsilon_{t}, \underline{\widetilde{\mathbf{y}}}_{t} \left(\underline{\rho}_{t}^{T} \left(\varepsilon^{t} \right) \right), \underline{\widetilde{\mathbf{x}}}^{t} \left(\underline{\rho}_{t}^{T} \left(\varepsilon^{t} \right) \right) - \underline{\mathbf{p}} \left(\underline{\rho}^{\tau T} \left(\varepsilon^{T} \right) \right) |\varepsilon_{0} \right] \\ - E_{\varepsilon^{T}} \left[\sum_{t=0}^{T} \omega_{t} \left(\varepsilon_{t}, \underline{\widetilde{\mathbf{y}}}_{t} \left(\underline{\rho}_{t} \left(\varepsilon^{t} \right) \right), \underline{\widetilde{\mathbf{x}}}^{t} \left(\underline{\rho}_{t} \left(\varepsilon^{t} \right) \right) - \underline{\mathbf{p}} \left(\underline{\rho}^{T} \left(\varepsilon^{T} \right) \right) |\varepsilon_{0} \right] . \end{split}$$

By the inductive hypothesis and the Lipschitz continuity of payoffs, for all $\delta, \tilde{\delta} > 0$, there are payment rules so that

$$|E_{\varepsilon^{T}}\left[\sum_{t=0}^{T}\omega_{t}\left(\varepsilon_{t},\underline{\mathbf{y}}_{t}\left(\underline{\rho}_{t}^{\tau}\left(\varepsilon^{t}\right)\right),\widetilde{\mathbf{x}}^{t}\left(\underline{\rho}_{t}^{\tau}\left(\varepsilon^{t}\right)\right)-\underline{\mathbf{p}}\left(\underline{\rho}^{\tau T}\left(\varepsilon^{T}\right)\right)|\varepsilon_{0}\right]$$

$$-E_{\varepsilon^{T}}\left[\sum_{t=0}^{T}\omega_{t}\left(\varepsilon_{t},\underline{\mathbf{y}}_{t}\left(\underline{\rho}_{t}\left(\varepsilon^{t}\right)\right),\widetilde{\mathbf{x}}^{t}\left(\underline{\rho}_{t}(\varepsilon^{t})\right)-\underline{\mathbf{p}}\left(\underline{\rho}^{T}\left(\varepsilon^{T}\right)\right)|\varepsilon_{0}\right]|\leq\delta,$$

$$E_{\varepsilon^{\tau+1}}\left[\left\|\begin{array}{c}\left(\underline{\widetilde{\mathbf{y}}}_{\tau+1}\left(\underline{\rho}_{\tau+1}\left(\varepsilon^{\tau+1}\right)\right),\widetilde{\mathbf{x}}^{\tau+1}\left(\underline{\rho}_{\tau+1}\left(\varepsilon^{\tau+1}\right)\right)\right)\\-\left(\underline{\widetilde{\mathbf{y}}}_{\tau+1}\left(\underline{\rho}_{t}^{\tau}\left(\varepsilon^{\tau+1}\right)\right),\widetilde{\mathbf{x}}^{\tau+1}\left(\underline{\rho}_{\tau+1}^{\tau}\left(\varepsilon^{\tau+1}\right)\right)\right)\right\|^{\sigma}\right]\leq\frac{\widetilde{m}\delta}{2K}$$

$$(22)$$

for $\sigma=1,2$ and (19) is satisfied for all $t\leq\tau.$ Therefore, by (21),

$$E_{\varepsilon^{T}}\left[\sum_{t=0}^{T}\omega_{t}\left(\varepsilon_{t},\underline{\widetilde{\mathbf{y}}}_{t}\left(\varepsilon^{t}\right),\underline{\widetilde{\mathbf{x}}}^{t}\left(\varepsilon^{t}\right)\right)-\underline{\mathbf{p}}\left(\varepsilon^{T}\right)|\varepsilon_{0}\right]$$
$$-E_{\varepsilon^{T}}\left[\sum_{t=0}^{T}\omega_{t}\left(\varepsilon_{t},\underline{\widetilde{\mathbf{y}}}_{t}\left(\underline{\rho}_{t}\left(\varepsilon^{t}\right)\right),\underline{\widetilde{\mathbf{x}}}^{t}\left(\underline{\rho}_{t}\left(\varepsilon^{t}\right)\right)-\underline{\mathbf{p}}\left(\underline{\rho}^{T}\left(\varepsilon^{T}\right)\right)|\varepsilon_{0}\right]$$
$$\geq E_{\varepsilon^{T}}\left[\sum_{t=0}^{T}\omega_{t}\left(\varepsilon_{t},\underline{\widetilde{\mathbf{y}}}_{t}\left(\varepsilon^{t}\right),\underline{\widetilde{\mathbf{x}}}^{t}\left(\varepsilon^{t}\right)\right)-\underline{\mathbf{p}}\left(\varepsilon^{T}\right)|\varepsilon_{0}\right]$$
$$-E_{\varepsilon^{T}}\left[\sum_{t=0}^{T}\omega_{t}\left(\varepsilon_{t},\underline{\widetilde{\mathbf{y}}}_{t}\left(\varepsilon^{t}\right),\underline{\widetilde{\mathbf{x}}}^{t}\left(\varepsilon^{t}\right)\right)-\underline{\mathbf{p}}\left(\underline{\rho}^{\tau}_{t}\left(\varepsilon^{T}\right)\right)|\varepsilon_{0}\right]-\delta$$

Note that according to the deviation strategy $\left\{\underline{\rho}_{t}^{\tau}\right\}_{t}$, the first deviation occurs in period $\tau + 1$. Therefore, we appeal to (14) once again and using the same arguments leading to (20) we conclude that

$$\begin{split} E_{\varepsilon^{T}} \left[\sum_{t=0}^{T} \omega_{t} \left(\varepsilon_{t}, \underline{\widetilde{\mathbf{y}}}_{t} \left(\varepsilon^{t} \right), \underline{\widetilde{\mathbf{x}}}^{t} \left(\varepsilon^{t} \right) \right) - \underline{\mathbf{p}} \left(\varepsilon^{T} \right) |\varepsilon_{0} \right] \\ &- E_{\varepsilon^{T}} \left[\sum_{t=0}^{T} \omega_{t} \left(\varepsilon_{t}, \underline{\widetilde{\mathbf{y}}}_{t} \left(\underline{\rho}_{t}^{\tau} \left(\varepsilon^{t} \right) \right), \underline{\widetilde{\mathbf{x}}}^{t} \left(\underline{\rho}_{t}^{\tau} \left(\varepsilon^{t} \right) \right) - \underline{\mathbf{p}} \left(\underline{\rho}^{\tau T} \left(\varepsilon^{T} \right) \right) |\varepsilon_{0} \right] \\ \geq \quad \widetilde{m} E_{\varepsilon^{\tau+1}} \left[\frac{\left\| \left(\underline{\widetilde{\mathbf{y}}}_{\tau+1} \left(\varepsilon^{\tau+1} \right), \underline{\widetilde{\mathbf{x}}}^{\tau+1} \left(\varepsilon^{\tau+1} \right) \right) - \left(\underline{\widetilde{\mathbf{y}}}_{\tau+1} \left(\underline{\rho}^{\tau}_{\tau+1} \left(\varepsilon^{\tau+1} \right) \right), \underline{\widetilde{\mathbf{x}}}^{\tau+1} \left(\underline{\rho}^{\tau}_{\tau+1} \left(\varepsilon^{\tau+1} \right) \right) \right) \right\|^{2}}{2K} \right] \\ \geq \quad \widetilde{m} E_{\varepsilon^{\tau+1}} \left[\frac{\left\| \left(\underline{\widetilde{\mathbf{y}}}_{\tau+1} \left(\varepsilon^{\tau+1} \right), \underline{\widetilde{\mathbf{x}}}^{\tau+1} \left(\varepsilon^{\tau+1} \right) \right) - \left(\underline{\widetilde{\mathbf{y}}}_{\tau+1} \left(\underline{\rho}_{\tau+1} \left(\varepsilon^{\tau+1} \right) \right), \underline{\widetilde{\mathbf{x}}}^{\tau+1} \left(\underline{\rho}_{\tau+1} \left(\varepsilon^{\tau+1} \right) \right) \right) \right\|^{2}}{2K} \right] - 3\delta_{\tau} \end{split}$$

where the last inequality follows from (22) and (18). Hence, in order to guarantee that the deviation strategy $(\rho_t)_t$ is profitable, we need that

$$E_{\varepsilon^{\tau+1}}\left[\left\|\left(\underline{\widetilde{\mathbf{y}}}_{\tau+1}\left(\varepsilon^{\tau+1}\right),\underline{\widetilde{\mathbf{x}}}^{\tau+1}\left(\varepsilon^{\tau+1}\right)\right)-\left(\underline{\widetilde{\mathbf{y}}}_{\tau+1}\left(\underline{\rho}_{\tau+1}\left(\varepsilon^{\tau+1}\right)\right),\underline{\widetilde{\mathbf{x}}}^{\tau+1}\left(\underline{\rho}_{\tau+1}\left(\varepsilon^{\tau+1}\right)\right)\right)\right\|^{2}\right] \leq \frac{2K(5\delta)}{\widetilde{m}}$$

So, choosing $\delta = \widetilde{m}\widetilde{\delta}/(10K)$ yields the claim in (19) for $\tau + 1$.

By the proof of Proposition 4, the payment rule can be defined so that

$$E_{\varepsilon^{T}} \sum_{t=0}^{T} \left\| f_{t} \left(\rho_{t} \left(\psi^{t} \left(\varepsilon^{t} \right) \right), \alpha_{t} \left(\psi^{t} \left(\varepsilon^{t} \right) \right) \right) - \underline{\mathbf{y}}_{t} \left(\underline{\rho}_{t} \left(\varepsilon^{t} \right) \right) \right\|^{\sigma} \leq \widetilde{\delta},$$
(23)

for $\sigma = 1, 2$. Define $(\overline{\mathbf{x}}_t(\psi_t(\varepsilon^t)), \overline{\mathbf{a}}_t(\psi_t(\varepsilon^t))) \equiv (\widetilde{\mathbf{x}}_t(\underline{\rho}_t(\varepsilon^t)), \widetilde{\mathbf{a}}_t(\underline{\rho}_t(\varepsilon^t)))$ for all t = 0, ..., T. Then

$$\begin{aligned} & \left\| \left(\widetilde{\mathbf{y}}_{t}\left(\theta^{t}\right), \widetilde{\mathbf{x}}_{t}\left(\theta^{t}\right) \right) - \left(\overline{\mathbf{y}}_{t}\left(\theta^{t}\right), \overline{\mathbf{x}}_{t}\left(\theta^{t}\right) \right) \right\| \\ & \leq \left\| \left(\underline{\mathbf{y}}_{t}\left(\varepsilon^{t}\right), \widetilde{\underline{\mathbf{x}}}_{t}\left(\varepsilon^{t}\right) \right) - \left(\underline{\mathbf{y}}_{t}\left(\underline{\rho}_{t}\left(\varepsilon^{t}\right) \right), \widetilde{\underline{\mathbf{x}}}_{t}\left(\underline{\rho}_{t}\left(\varepsilon^{t}\right) \right) \right) \right\| \\ & + \left\| f_{t}\left(\rho_{t}\left(\psi^{t}\left(\varepsilon^{t}\right)\right), \alpha_{t}\left(\psi^{t}\left(\varepsilon^{t}\right)\right) \right) - \underline{\mathbf{y}}_{t}\left(\underline{\rho}_{t}\left(\varepsilon^{t}\right) \right) \right\|. \end{aligned}$$

Notice that summing up the inequalities in (19) for t = 0, ..., T yield

$$E_{\varepsilon^{T}}\sum_{t=0}^{T}\left\|\left(\underline{\mathbf{y}}_{t}\left(\varepsilon^{t}\right),\underline{\widetilde{\mathbf{x}}}_{t}\left(\varepsilon^{t}\right)\right)-\left(\underline{\mathbf{y}}_{t}\left(\underline{\rho}_{t}\left(\varepsilon^{t}\right)\right),\underline{\widetilde{\mathbf{x}}}_{t}\left(\underline{\rho}_{t}\left(\varepsilon^{t}\right)\right)\right)\right\|^{2}<\widetilde{\delta}.$$

By (18), the previous inequality and (23), we conclude that

$$E_{\theta^{T}} \sum_{t=0}^{T} \left\| \left(\widetilde{\mathbf{y}}_{t} \left(\theta^{t} \right), \widetilde{\mathbf{x}}_{t} \left(\theta^{t} \right) \right) - \left(\overline{\mathbf{y}}_{t} \left(\theta^{t} \right), \overline{\mathbf{x}}_{t} \left(\theta^{t} \right) \right) \right\|^{2} \leq 4 \widetilde{\delta}.$$

Therefore, the allocation $(\overline{\mathbf{x}}_t, \overline{\mathbf{a}}_t)_t$ satisfy desired inequality in the statement of Proposition 5. Finally, note that $(\overline{\mathbf{x}}_t, \overline{\mathbf{a}}_t)_t$ is implementable because it results from the agent's optimal deviation strategy in the mechanism $(\widetilde{\mathbf{x}}^T, \widetilde{\mathbf{a}}^T, \widetilde{\mathbf{p}})$. \Box

References

Eső, Péter and Balázs Szentes (2007): "Optimal Information Disclosure in Auctions and the Handicap Auction," *Review of Economic Studies* 74(3), 705-731.