International environmental cooperation and dynamic stock effects[*]

Marita Laukkanen[†]   Yulia Pavlova[‡]

January 25, 2017

Abstract

Many key environmental concerns today involve strategic interactions among sovereign countries. While most transboundary pollution problems involve stock pollutants, much of the literature on international environmental agreements (IEAs) uses static models. Such models ignore the fact that changes in the stock of pollution may affect the incentives of countries to join an IEA.

In this paper we analyze effect of pollutant stock dynamics on IEA membership assuming that at any time period each country makes a decision whether to participate in an IEA or not, and depending on the outcome on these decisions, each country adjusts its emission policy according to the feedback Nash equilibrium. Our objective is to find equilibrium size of an IEA, which withstands possible one-shot deviations. The model is applied to a numerical calibration reflecting the climate policy problem. We show that optimal emission policy is determined as a function of the current pollution level and equilibrium size of the IEA membership strongly depends on whether a potential deviator may or may not be allowed to rejoin the agreement after the deviation: in case when a deviator can return to the agreement at the next stage following her deviation, the agreement at the steady state contains only two countries, however if a deviator can not return, the agreement size can be very large.

JEL codes: Q50, Q53, Q54, Q58

Keywords: International agreements, game theory, stock pollution

1　Introduction

Many key environmental concerns today involve strategic interactions among sovereign countries. Some important problems are global in nature: climate change, depletion of the ozone layer, and loss of biodiversity call for policy coordination at the global level. Marine pollution, eutrophication and depletion of fish stocks and marine ecosystem services may be addressed at a more regional level but still call for measures to be undertaken in many countries. In practice global and international environmental cooperation has proven challenging. What makes reaching an international agreement particularly difficult is that countries cannot be forced to participate and comply with an agreement but must do so voluntarily.

There is an extensive economics literature on participation in international environmental agreements (IEAs), starting with Carraro and Siniscalco (1993) and Barrett (1994). This literature builds on a static two-stage non-cooperative game, developed by d'Aspremont et al. (1983) in cartel theory. In the first stage, countries decide whether they want to particiate in the agreement or not. In the second stage, a coalition of IEA members and the individual outsiders choose their production and emission levels. The subgame perfect Nash equilibrium of the resulting two-stage game is usually interpreted in terms of internal and external stability of the agreement: no signatory has an incentive to leave the agreement and no outsider has an incentive to join, which is the Nash equilibrium of the first stage given the equilibrium in the second stage. The basic static model yields the pessimistic conclusion that free-rider incentives are stronger than incentives to cooperate and that the size of the stable coalition is small, regardless of the total number of countries.

While important insights have been developed in this literature, it largely fails to recognize that most pollutants accumulate in the environment over time and that environmental harm is caused by the accumulated pollution rather than current emissions (e.g., concentrations of greenhouse gases in the atmosphere or acidification of soils). The incentives to participate in an IEA may then change over time, as the pollution stock changes. There are only a few published papers that include both the participation decision and stock dynamics as elements of the game. Rubio and Casino (2005) adapt the model developed in Carraro and Siniscalco (1993) and Barrett (1994) to a dynamic framework but do not really address the impact of the pollution stock on the prospects of cooperation - they consider an open loop solution where countries commit to their initial period membership status and emission paths for an infinite length of time. Rubio and Ulph (2007) extend the model in Rubio and Casino (2005) to the case where countries commit to their current membership

decision for the duration of one period only, and are thus able to update their membership status as the stock level changes. In their model, the per period payoffs consist of a linear benefit from emissions and a quadratic damage from the stock of accumulated pollution. Membership of the IEA is allowed to vary depending on the evolution of the stock. The paper concludes that the prospects for cooperation are not very bright: the size of the stable coalition decreases as the emission stock approaches its steady state.

While Rubio and Ulph (2007) is an important step towards analyzing countries' strategic incentives for participation in an environmental treaty, the model is of limited applicability due to a number of strong assumptions made in order to achieve analytical results. First, the model can only be applied to cases where the number of potential participants is large, on the order of at least 20-30 countries. This will be true for global problems such as climate change, but does not hold in the case of local pollutants and regional treaties, where number of participants is typically smaller. Second, the assumption of linear benefits from emissions is unlikely to find empirical support. Finally, the numerical illustrations presented in the paper concern the case where the current pollution stock is below the cooperative steady state, whereas in many real-world the opposite is more likely to hold. Models similar to the Rubio and Ulph (2007) one have been considered by Breton et al. (2010), who apply replicator dynamics to analyze a similar model, and Nkuiya (2012), who develops a continuous-time version of the Rubio and Ulph (2007) model, and treats the length of the period of commitment as a parameter that can take on any strictly positive value.

Battaglini and Harstad (2015) examine the hold-up problem related to underinvestment in green technologies in the context of an IEA and a stock pollutant. The paper considers a symmetric dynamic game where individual payoffs in each period are composed of the sum of disutility from reduced individual consumption, environmental damage and investments in green technology. The model has two decision variables, emission reduction (interpreted as reduced consumption) and investments, and two stock variables, an emission stock and a technology stock. Two IEA structures are compared: one where the IEA is a contract on both emissions and investments, and one where the IEA is a commitment on emissions only while investments are set individually, and assumed non-tradable. It is also assumed that countries can commit to participation (or non-participation) in the agreement for an endogenous length of time that depends on the coalition size. However, Battaglini and Harstad (2015) focus on the hold-up problem in different contracting environments, and do not allow the pollution stock to have an effect on IEA participation: They assume that the damage from pollution is linear, which has the consequence that equilibrium emissions are

3

independent of the pollution stock.

In this paper, we extend the Rubio and Ulph (2007) approach to the more realistic work-horse model of IEAs: countries' per period payoffs are identical and equal to quadratic benefit from production minus quadratic environmental damage (a one-to-one relationship is assumed between production and emission). In contrast to the Battaglini and Harstad (2015) model, equilibrium emission decisions then depend on the current level of pollution. We focus on international cooperation on emission reductions, leaving the hold-up problem with dynamic stock effects for future study. Unlike the Rubio and Ulph (2007) model, no restrictions need to be placed on the number of participating countries, neither do we rely on random assignment rule for determining which countries become signatories. Instead, we consider a two-stage game over the entire time horizon of the game, rather than in each period and proceed by solving the emission game for any coalition size and stock level. The feedback Nash solution to the emission game determines the emissions of signatories and non-signatories as functions of the stock level with coalition size entering the control functions as a parameter. We then solve the membership game for any stock level and equilibrium pollution control strategies. Since we have assumed that the countries are identical, the identity of the countries in the coalition of IEA members is irrelevant. We are interested in characterizing the equilibrium number of countries that participate in the IEA at any stock level, and will henceworth ignore the identity of countries in the equilibrium coalition. The model is applied to a numerical calibration reflecting the climate policy problem. We show that equilibrium size of the IEA membership strongly depends on whether a potential deviator may or may not be allowed to rejoin the agreement after the deviation: in case when a deviator can return to the agreement at the next stage following her deviation, the agreement at the steady state contains only two countries, however if a deviator can not return, the agreement size can be very large.

The paper is organized as follows. In Section 2 we present the game-theoretic model with stock pollutant, as well as first best (full cooperation case) and no cooperation case, whether we derive optimal emission policies and steady state conditions. Section 3 describes approach to model IEAs with participation changing over time. As in the previous literature, we assume two stage game played at every period of the game: membership game (each country decides whether cooperate or not in the agreement in order to reduce emissions) and emission game (countries consequently adjust their emission behavior over time). Section 3.1 shows how we solve emission game and find steady state emission policies for any given size of the agreement (as feedback Nash equilibrium). In Section 3.2 we proceed with analysis of the number of signatories along the path to the steady state. First, we analyze countries'

incentives to participate in an IEA assuming that a signatory might deviate for one period and can rejoin the agreement in the following course of the game. Secondly, we assume 'infinite punishment' for the deviator: after one-shot deviation, it can never become an IEA member again. A stand-alone outsider is always permitted to join the agreement for one period. Subsection 3.3 presents steady state analysis of the membership game, first, under assumption of one-shot deviation, and, secondly, under infinite punishment. In Section 4 and 5 we proceed with numerical application and analysis of the game. Section 6 presents conclusive discussion.

2   Model and preliminaries

We consider N identical players, each of which emits a pollutant q_{it}, $i = 1, \ldots, N$ and $t = 1, 2, \ldots$, that damages a shared environmental resource. A one-to-one relationship is assumed between production and emissions so that $q_{it} \geq 0$ denotes both production and emissions. Environmental damage is a quadratic function of the emission stock z_t, $t = 0, 1, \ldots$. The per period net benefit to each player depends on her own production q_{it} and on the pollution stock z_t :

$$\Pi_{it}(q_{it}, z_t) = \alpha(\beta q_{it} - \frac{1}{2} q_{it}^2) - \frac{1}{2} \gamma z_t^2.$$

Later we refer to the benefit from emissions as $\pi_{it} = \alpha(\beta q_{it} - \frac{1}{2} q_{it}^2)$. The stock of pollution evolves according to

$$z_{t+1} = \rho z_t + \sum_{i=1}^{N} q_{it}, \tag{1}$$

where $\rho \geq 0$ is the carry-over of the pollution stock. The initial stock is given by z_0. The objective of each player is to choose the emissions q_{it}, $t = 1, 2, \ldots$, so as to maximize her expected net benefits over time. In the absence of environmental damage, production q_{it} would be equal to β, which we refer to as the business-as-usual emission level. Without loss of generality, we normalize the model so that $\beta = 1$. The marginal benefit and damage parameters α and γ are positive and identical among the players.

2.1   The first-best solution: full cooperation

Suppose now that all the N countries cooperate and coordinate their emissions to maximize the present value of joint welfare. We consider a utilitarian joint welfare function, where each country's individual welfare has weight 1. The per-period joint welfare then is $W_t(\cdot) = \sum_{i \in N} \Pi_{it}(q_{it}, z_t)$. The countries maximize the present value of joint welfare,

$$\sum_{t=1}^{\infty} \delta^{t-1} \sum_{i=1}^{N} \Pi_{it}(q_{it}, z_t),$$

subject to the state equation (1). We solve this problem by dynamic programming. The Bellman equation for the joint welfare maximization problem is

$$V(z_t) = \sum_{i=1}^{N} \Pi_{it}(q_{it}, z_t) + \delta V(z_{t+1}). \tag{2}$$

As the problem is linear-quadratic, the value function will be quadratic in z. We define the quadratic value function as $V(z) = \frac{1}{2} A_c z_t^2 + B_c z_t + C_c$. The first-order conditions for maximizing (2) are then straightforward to derive. The first-best emission levels are given by

$$q_{it}^c = \frac{\rho \delta A_c}{\alpha - \delta N A_c} z_t + \frac{\delta B_c + \alpha}{\alpha - \delta N A_c}. \tag{3}$$

The value function coefficients A_c, B_c and C_c can be solved by inserting $V(z) = \frac{1}{2} A_c z_t^2 + B_c z_t + C_c$ and the optimal emissions (3) into (2) and equating the coefficients.

The symmetric steady state, with $q_{it} = q^*$, is given by

$$\alpha(1 - q^*) + N\delta\lambda^* = 0$$
$$-\gamma z^* + \delta\rho\lambda^* - \lambda^* = 0$$
$$\rho z^* - z^* + Nq^* = 0$$

where λ^* is the steady state shadow value of the pollution stock.

2.2 No cooperation

Suppose instead that all countries choose their emission levels simultaneously and in a noncooperative way. In a Markov-perfect equilibrium country i assumes that its choice of q_{it} will not affect the future choice of $q_{j\tau}$ for any player j or time $\tau = t+1, \dots$. Furthermore, each country takes the other countries' current emissions as given. Each country is then maxizing

$$\sum_{t=1}^{\infty} \delta^{t-1} \Pi_{it}(q_{it}, z_t),$$

subject to the state equation $z_{t+1} = \rho z_t + q_{it} + \sum_{j \neq i} q_{jt}$. Again, we solve the problem using dynamic programming. The Bellman equation for the individual country's problem is

$$V_i(z_t) = \Pi_{it}(q_{it}, z_t) + \delta V_i(z_{t+1}). \tag{4}$$

The individual country's problem is also linear-quadratic, and the value function will be quadratic in z. We define the quadratic value function as $V_i(z) = \frac{1}{2}A_n z_t^2 + B_n z_t + C_n$. Again, the first-order conditions for maximizing (4) are straightforward to derive. The non-cooperative emission levels are given by

$$q_{it}^n = \frac{\rho \delta A_n}{\alpha - \delta N A_n} z_t + \frac{\delta B_n + \alpha}{\alpha - \delta N A_n}.$$

The coefficients A_n, B_n and C_n can again be solved in the standard way.

The symmetric steady state, with $q_{it} = q^o$, is given by

$$
\begin{aligned}
\alpha(1 - q^o) + \delta \lambda^o &= 0 \\
-\gamma z^o + \delta \rho \lambda^o - \lambda^o &= 0 \\
\rho z^o - z^o + N q^o &= 0
\end{aligned}
$$

where λ^o is the steady state shadow value of the pollution stock when the countries do not cooperate.

It is easy to show that for any $N > 1$ the steady-state stock and per-period emissions are lower in the cooperative equilibrium than in the non-cooperative equilibrium, and the discounted present value of net benefits per country is higher in the cooperative equilibrium thatn in the non-cooperative equilibrium. The non-coopearative equilibrium coincides with the first-best only if $N = 1$.

3 IEA with participation changing over time

We now turn to the case of partial cooperation. We examine the formation of an infinite sequence of IEAs, where in each period some countries participate in the IEA while some countries choose to stay out. The structure and the timing of the game are as follows: At the beginning of each period t, the N countries observe the size of the pollution stock z_t. Countries can condition both their IEA membership and emissions at time t on the pollution stock at time t. Countries may adjust their strategy at each point in time. In particular, in each period, countries are free to join or leave the agreement. This structure captures the fact that countries enter IEAs voluntarily and may at any time abandon the process. Canada, for example, ratified the Kyoto protocol in 2002 but withdraw from the agreement in 2011. The model of IEA formation is a dynamic version of the work-horse model of self-enforcing IEAs developed by Carraro and Siniscalco (1993) and Barrett (1994).

The emission game is defined by a triple that consists of the set of players, the players' strategy space, and the players's payoff vectors $\langle \mathcal{N}, \{q_{it}\}_{i=1,t=1}^{N,\infty}, \{\Pi_{it}\}_{i=1,t=1}^{N,\infty} \rangle$ and a set of

rules. The fact that countries may adjust their emissions at any time implies that we look for the feedback Nash or Markov Perfect equilibrium of the emission game. As the payoffs do not depend on time t, the solution to the emission game will be stationary, meaning that the equilibrium emissions will only be a function of the pollution stock and not a function of time.

A country may decide whether or not to cooperate with other countries in order to reduce total emissions at any time t. That is, countries may also adjust their IEA membership at any time. The decision regarding whether or not to cooperate is the outcome of a 'metagame', called a membership game in the IEA literature. In the membership game, each country choose between the cooperative and the non-cooperative strategy (IEA membership versus fringe) by anticipating the outcomes of the related emission game.

The feedback Nash or Markov perfect equilibrium of the emission game can be found by solving the dynamic programming equations in the value functions of the countries. At each time t, IEA members choose emissions to maximize the expected total net benefits of all signatory countries, taking as given the behavior of non-signatories. The non-signatories in turn simultaneously and independently choose emissions to maximize their own, individual expected net benefits. At this stage, the size of the IEA is considered fixed. However, the value functions determined by the feedback Nash equilibrium will depend on the size of the IEA: for each IEA size, there will be value functions and **feedback controls** for signatories and non-signatories that determine emissions as a function of the pollution stock.

Given emission game equilibrium, the membership game equilibrium outcome will be determined through internal and external stability of the agreement: no signatory has an incentive to leave the agreement and no outsider has an incentive to join (the Nash equilibrium of the membership game). In particular, the equilibrium IEA size at stock level z_t is a number of signatories n such that no signatory country would wish to switch to being a non-signatory (internal stability), and no non-signatory would wish to switch to being a signatory (external stability).

Differing from Rubio and Ulph (2007), we consider a two-stage game over the entire time horizon of the game, rather than in each period, as has been done by Rubio and Ulph (2007). That is, we proceed by solving the emission game for any n and z_t. The feedback Nash/MPE solution to the emission game determines the emissions of signatories and non-signatories as functions of the stock level z_t, with coalition size n entering the control functions as a parameter. We then solve the membership game for any z_t, given the feedback Nash emissions at stock z_t for each n. Timing of the game is presented in Fig. 1.

Since we have assumed that the countries are identical, the identity of the countries

in the coalition S of IEA members is irrelevant. That is, if we have an equilibrium with coalition S, we will have an equilibrium with any other coalition $S' \neq S$ with $|S'| = |S|$. We are interested in characterizing the equilibrium *number* of countries that participate in the IEA at any stock level z_t, and will henceworth ignore the identity of countries in the equilibrium coalition.
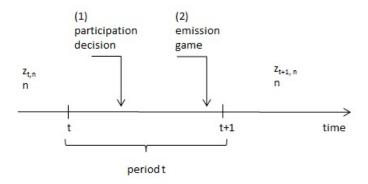

Figure 1: Timing of the game

In the following course of analysis, we shall use the following notation. The state variable in the partial membership game is

$$z_t \in [0, \infty),$$

and the action variables are

$$q_{i,t}^s \in [0, \infty),$$
$$q_{j,t}^f \in [0, \infty)$$

where $q_{i,t}^s$ are a signatory country's emissions in period t and $q_{j,t}^f$ a non-signatory country's emissions. The state transition function for the pollution stock then is

$$z_{t+1} = \rho z_t + \sum_{i=1}^{n} q_{i,t}^s + \sum_{j=1}^{N-n} q_{j,t}^f,$$

where ρ is carry-over and N the number of countries in total.

3.1 Emission game

Suppose that there are are n members in the IEA. With n given, the **group of signatories and each non-signatory country** face a standard deterministic linear-quadratic game. Signatories' and non-signatories' control rules are linear in the stock, and the value functions are quadratic. Thus, we can use standard methods to find **Nash feedback controls**. Let

$V_n^s(z_t)$ and $V_n^f(z_t)$ denote the value functions that describe the continuation values for signatories and non-signatories given a pollution stock z_t. Here, $V_n^s(z_t)$ is the joint value function for *all* signatory countries while $V_n^f(z_t)$ is the value function for an individual non-signatory country j. The value function coefficients are different under each value of n.

The Bellman equation for a non-signatory (or fringe) country, given n, is

$$V_n^f(z_t) = \max\left\{\pi\left(q_t^f\right) - \gamma\frac{1}{2}z_t^2 + \delta V_n^f(z_{t+1})\right\} \tag{5}$$

The Bellman equation for signatories (which optimize a group) is:

$$V_n^s(z_t) = \max\left\{n\pi\left(q_t^s\right) - n\gamma\frac{1}{2}z_t^2 + \delta V_n^s(z_{t+1})\right\} \tag{6}$$

Given (5) and (6), the first-order condition for the non-signatories' problem is

$$\frac{\partial\pi\left(q_t^f\right)}{\partial q_t^f} + \delta\frac{\partial V_n^f(z_{t+1})}{\partial z_{t+1}} = 0,$$

as each fringe country only accounts for its own emissions and takes the other countries emissions as given.

The first-order condition for the signatories' problem is

$$\frac{\partial\pi\left(q_t^s\right)}{\partial q_t^s} + \delta\frac{V_n^s(z_{t+1})}{\partial z_{t+1}} = 0.$$

The signatories maximize their joint value, accounting for the effect of all signatories' emissions on the pollution stock.

In Appendix A [to be completed later, Mathematica file] we show that the optimal values of q_t^s and q_t^f are linear in the pollution stock, and the value functions $V_n^f(z_t)$ and $V_n^s(z_t)$ are quadratic in the stock z_{t+1}:

$$q_t^s = \frac{a^2 + \delta^2(N-n)(A_sB_n - A_nB_s) + a\delta(B_s + (N-n)(A_n - A_s)) + a\delta\rho A_sz}{a(a - \delta NA_n + \delta n(A_n - A_s))} \tag{7}$$

$$q_t^f = \frac{a^2 + \delta^2n(A_nB_s - A_sB_n) + a\delta(B_n + n(A_n - A_s)) + a\delta\rho A_nz}{a(a - \delta NA_n + \delta n(A_n - A_s))} \tag{8}$$

$$V_n^f(z) = \frac{1}{2}A_n^fz^2 + B_n^fz + C_n^f,$$
$$V_n^s(z) = \frac{1}{2}A_n^sz^2 + B_n^sz + C_n^s.$$

The coefficients of these value functions satisfy

$$A_n^s = ... \tag{9}$$

$$B_n^s = ... \tag{10}$$

$$C_n^s = ... \tag{11}$$

$$A_n^f = ... \tag{12}$$

$$B_n^f = ... \tag{13}$$

$$C_n^f = ... \tag{14}$$

Given that the value functions are bounded above (for $\delta < 1$) and that they are quadratic, it must be the case that $A_n^f \leq 0$ and $A_n^s \leq 0$. Equation (12) is quadratic in A_n^f and equation (9) in A_n^s. We first solve the **system of equations** (12) and (9) to obtain the unique negative roots of this system of equations. We then solve **the linear systems of equations** (13) and (10) and (14) and (11) recursively. **Given the values of A_n^f to C_n^s** we use equations (7) and (8) to obtain signatories' and non-signatories' optimal levels of emissions at stock level z_t under partial cooperation with any n IEA members.

3.1.1 Steady state emissions

Let the steady state stock corresponding to IEA size n be given by z_n^*. In the steady state, the emissions q_n^{f*} and q_n^{s*} have to satisfy the following Euler and state stationarity conditions:

$$\frac{\partial \pi \left(q_n^{f*} \right)}{\partial q^f} + \delta \frac{\partial V_n^f \left(z_n^* \right)}{\partial z} = 0$$

$$\frac{\partial \pi \left(q_n^{s*} \right)}{\partial q^s} + \delta \frac{V_n^s \left(z_n^* \right)}{\partial z} = 0$$

$$z^* = \rho z^* + n q_n^{s*} + (N - n) q_n^{f*}$$

3.2 Number of signatories (membership game) along the path to steady state

The value functions $V_n^f(z)$ and $V_n^s(z)$ describe the value of **the optimal feedback solution** for each non-signatory country and the group of signatories at stock level z, for a given IEA size n. We denote the continuation values for a non-signatory country and an individual signatory country that maintain their period $t-1$ membership status in period t by $W_n^f(z_t)$ and $W_n^s(z_t)$. With $q_n^f(z_t)$ and $q_n^s(z_t)$ denoting the optimal emissions of non-signatories and signatories, these continuation values are as follows:

$$W_n^f(z_t) = V_n^f(z_t) = \pi\left[q_n^f(z_t)\right] - \gamma\frac{1}{2}(z_t)^2 + \delta V_n^f(z_{t+1})$$

and, since $V_n^s(z_t)$ is the joint value function for all signatories,

$$W_n^s(z_t) = \frac{1}{n}V_n^s(z_t) = \pi\left[q_n^s(z_t)\right] - \gamma\frac{1}{2}(z_t)^2 + \frac{1}{n}\delta V_n^s(z_{t+1}).$$

We next consider the internal and external stability of the agreement. We define internal and external stability as follows: an agreement is internally stable at stock level z_t if no signatory wishes to leave the agreement, and externally stable at stock level z_t if no non-signatory wishes to join the agreement at time t. If a country that is a member in the IEA withdraws from the agreement at time t, it will emit as a non-signatory country in period t. Similarly, if a non-member country joins the agreement in period t, it will emit as a signatory in period t.

3.2.1 Stability check 1: no punishment

We first consider stability when the rules of the agreement allow countries to leave and then rejoin the agreement at any time. A signatory country withdrawing from the agreement is then not punished in any way. To check whether such an agreement is stable with n members at stock size z_t, we utilize the one-time deviation principle: the agreement with n members is stable at stock level z_t if there are no profitable one-time deviations for signatories or non-signatories. That is, an agreement with n members is internally stable at stock level z_t if there are no profitable deviations where a signatory country i would exit the agreement in period t but rejoin it in period $t+1$. An agreement with n members is externally stable at stock level z_t if there are no profitable deviations where a non-signatory country j would join the agreement in period t and exit it in period $t+1$. For each pollution level z_t we then take the stable size of the agreement to be the largest n that is internally and externally stable.

Suppose that a signatory country i exits the agreement in period t. Country i then emits as a non-signatory in period t but returns to being a signatory in period $t+1$. The pollution stock in period $t+1$ obtains the value

$$z_{t+1,n-1} = \rho z_t + (n-1) q_{n-1}^s (z_t) + [N - (n-1)] q_{n-1}^f (z_t), \tag{15}$$

where $q_{n-1}^s (z_t)$ and $q_{n-1}^f (z_t)$ are signatories' and non-signatories' optimal emissions with the period t stock z_t and $n-1$ signatories. With a one-period deviation, the deviating signatory's period $t+1$ value function is the one for signatories, but the stock in period $t+1$ is different from the stock along the path with n signatories. The continuation payoff to a signatory country deviating from the agreement then is

$$W_{n-1}^{d-} (z_t) = \pi \left[q_{n-1}^f (z_t) \right] - \gamma \frac{1}{2} (z_t)^2 + \frac{1}{n} \delta V_n^s (z_{t+1,n-1}) .$$

Consider next the case where a non-signatory country j enters the agreement in period t but exits it again in period $t+1$. Country j then emits as a signatory in period t but returns to emitting as a non-signatory in period $t+1$. The pollution stock in period $t+1$ is given by

$$z_{t+1,n+1} = \rho z_t + (n+1) q_{n+1}^s (z_t) + [N - (n+1)] q_{n+1}^f (z_t), \tag{16}$$

where $q_{n+1}^s (z_t)$ and $q_{n+1}^f (z_t)$ are signatories' and non-signatories' optimal emissions with the period t stock z_t and $n+1$ signatories. Country j's period $t+1$ value function is the one for a non-signatory country.

The continuation payoff to a non-signatory country entering the agreement then is

$$W_{n+1}^{d+} (z_t) = \pi \left[q_{n+1}^s (z_t) \right] - \gamma \frac{1}{2} (z_t)^2 + \delta V_n^f (z_{t+1,n+1}) . \tag{17}$$

The internal and external stability conditions at any IEA size n and stock level z_t can be written as follows:

$$W_n^s (z_t) \geq W_{n-1}^{d-} (z_t) \tag{18}$$

$$W_n^f (z_t) \geq W_{n+1}^{d+} (z_t) \tag{19}$$

The internal stability condition (18) states that no signatory has an incentive to leave the coalition, and the external stability condition (19) states that no non-signatory has an incentive to join the coalition when the coalition has n members and the pollution stock is z_t.

The internal stability condition can be written out as

$$J (z_t) = W_n^s (z_t) - W_{n-1}^{d-} (z_t) = \pi [q_n^s (z_t)] - \pi \left[q_{n-1}^f (z_t) \right] \tag{20}$$

$$+ \delta \frac{1}{n} [V_n^s (z_t) - V_n^s (z_{t+1,n-1})] \geq 0.$$

The external stability condition in turn equals

$$W_{n+1}^{d+}(z_t) - W_n^f(z_t) = \pi\left[q_{n+1}^s(z_t)\right] - \pi\left[q_n^f(z_t)\right] \tag{21}$$
$$+\delta\left[V_n^f(z_{t+1,n+1}) - V_n^f(z_t)\right] \leq 0.$$

The stable size of the agreement at any stock level z_t is the largest n that satisfies the internal and external stability conditions (20) and (21) at z_t.

The payoffs from deviating, $W_n^s(z_t) - W_{n-1}^{d-}(z_t)$ and $W_{n+1}^{d+}(z_t) - W_n^f(z_t)$, are quadratic in z_t (as shown in Appendix). Suppose for now that $W_n^s(z_t) - W_{n-1}^{d-}(z_t)$ is concave in z_t and $W_{n+1}^{d+}(z_t) - W_n^f(z_t)$ convex in z_t[1].

Then, if $W_n^s(z_t) - W_{n-1}^{d-}(z_t) = 0$ has roots z_1 and z_2, with $z_1 \leq z_2$, an agreement with n members is internally stable for all stock levels $z_1 \leq z_t \leq z_2$. If $W_n^s(z_t) - W_{n-1}^{d-}(z_t) = 0$ only has one root z_1, an agreement with n members is internally stable only at stock level z_1. If $W_n^s(z_t) - W_{n-1}^{d-}(z_t) = 0$ has no roots, there are no z_t for which an agreement with n members is internally stable.

Similarly, if $W_{n+1}^{d+}(z_t) - W_n^f(z_t) = 0$ has roots z_3 and z_4, with $z_3 \leq z_4$, an agreement with n members is externally stable for all stock levels $z_3 \leq z_t \leq z_4$. If $W_n^s(z_t) - W_{n-1}^{d-}(z_t) = 0$ only has one root z_3, an agreement with n members is externally stable only at stock level z_3. If $W_n^s(z_t) - W_{n-1}^{d-}(z_t) = 0$ has no roots, there are no z_t for which an agreement with n members is externally stable.

3.2.2 Stability check 2: infinite punishment

We next consider stability when a signatory country that exits the agreement is not allowed to rejoin. This amounts to an infinite punishment to a signatory country deviating from the agreement. If there are no profitable deviations at stock level z_t where a signatory country i leaves the agreement in period t to remain a non-signatory thereafter, membership size n is internally stable at z_t. As it does not seem plausible that a non-signatory country entering the agreement could be prevented from ever leaving the agreement, in terms of external stability we maintain the same rule as in the previous section - a non-signatory joining the agreement is free to leave it in the following period. Hence, external stability still corresponds to there being no profitable one-time deviations where a non-signatory country j joins the agreement in period t and exits the agreement in period $t+1$.

Suppose a signatory country i deviates from the agreement. Country i then emits as a non-signatory from the current period onwards, and remains a non-signatory up to infinity.

---

[1]To be developed analytically, proposition. Otherwise, verified numerically.

Thus, its value function in the following period is the one for a non-signatory country. The stock in period $t+1$ is again different from the stock along the path with n signatories, obtaining the value given by Equation (15).The continuation payoff to a signatory country deviating from the agreement then is

$$W_{n-1}^{d-;\infty}(z_t) = \pi \left[q_{n-1}^f(z_t) \right] - \gamma \frac{1}{2} (z_t)^2 + \delta V_{n-1}^f(z_{t+1,n-1}).$$

A one-time deviation by a non-signatory country j in period t is defined as in the previous section. Country j emits as a signatory in the current period but returns to emitting as a non-signatory in the following period. The stock in period $t+1$ obtains the value in Equation (16) and the continuation payoff to a non-signatory country entering the agreement the value in Equation (17).

The internal and external stability conditions in the steady state can be written as follows:

$$W_n^s(z_t) \geq W_{n-1}^{d-\infty}(z_t) \tag{22}$$

$$W_n^f(z_t) \geq W_{n+1}^{d+}(z_t) \tag{23}$$

As before, the internal stability condition (22) states that no signatory has an incentive to leave the agreement, and the external stability condition (23) states that no non-signatory has an incentive to join the agreement.

The internal stability condition can be written out as

$$W_n^s(z_t) - W_{n-1}^{d-\infty}(z_t) = \pi [q_n^s(z_t)] - \pi \left[q_{n-1}^f(z_t) \right] \tag{24}$$
$$+ \delta \left[\frac{1}{n} V_n^s(z_t) - V_n^f(z_{t+1,n-1}) \right] \geq 0.$$

The external stability condition is again given by Equation (21).

3.3   Steady state membership

Let n denote a candidate IEA size whose stability at the steady state we are checking, and z_n^* the corresponding steady state stock. Let $W_n^s(z_n^*)$ and $W_n^f(z_n^*)$ be the continuation values for an individual signatory and non-signatory country when each country maintains its steady state membership status. With q_n^{s*} and q_n^{f*} denoting the optimal emissions of signatories and non-signatories with IEA size n and steady state stock z_n^*, the continuation values $W_n^s(z_n^*)$ and $W_n^f(z_n^*)$ at the steady state are as follows:

$$W_n^f(z_n^*) = V_n^f(z_n^*) = \pi \left(q_n^{f*} \right) - \gamma \frac{1}{2} (z_n^*)^2 + \delta V_n^f(z_n^*)$$

and, since $V_n^s(z_t)$ is the joint value function for all signatories,

$$W_n^s(z_n^*) = \frac{1}{n} V_n^s(z_n^*) = \pi(q_n^{s*}) - \gamma \frac{1}{2}(z_n^*)^2 + \frac{1}{n}\delta V_n^s(z_n^*).$$

3.3.1 Stability check 1: no punishment

Again, we first consider stability when countries deviating from the agreement are not punished, and examine whether there are profitable one-time deviations for signatories or non-signatories. Suppose a signatory country i deviates from the agreement. The next period stock will differ from the steady state value z_n^*, as the change in membership from n to $n-1$ in period t affects optimal emissions $q_t^f(z_n^*)$ and $q_t^s(z_n^*)$. The stock now obtains the value

$$z_{n-1}' = \rho z_n^* + (n-1) q_{n-1}^s(z_n^*) + [N - (n-1)] q_{n-1}^f(z_n^*), \tag{25}$$

where $q_{n-1}^s(z_n^*)$ and $q_{n-1}^f(z_n^*)$ are signatories' and non-signatories' optimal emissions with the period t stock z_n^* and $n-1$ signatories. The continuation payoff to a signatory country deviating from the agreement then is

$$W_{n-1}^{d-}(z_n^*) = \pi\left(q_{n-1}^f(z_n^*)\right) - \gamma \frac{1}{2}(z_n^*)^2 + \frac{1}{n}\delta V_n^s(z_{n-1}').$$

Suppose next that a non-signatory country j deviates in period t. The stock in period $t+1$ again differs from its steady state value z_n^*, obtaining the value

$$z_{n+1}' = \rho z_n^* + (n+1) q_{n+1}^s(z_n^*) + [N - (n+1)] q_{n+1}^f(z_n^*), \tag{26}$$

where $q_{n+1}^s(z_n^*)$ and $q_{n+1}^f(z_n^*)$ are signatories' and non-signatories' optimal emissions with the period t stock z_n^* and $n+1$ signatories. The continuation payoff to the non-signatory country entering the agreement for one period then is

$$W_{n+1}^{d+}(z_n^*) = \pi\left(q_{n+1}^s(z_n^*)\right) - \gamma \frac{1}{2}(z_n^*)^2 + \delta V_n^f(z_{n+1}'). \tag{27}$$

The internal and external stability conditions in the steady state can be written as follows:

$$W_n^s(z_n^*) \geq W_{n-1}^{d-}(z_n^*) \tag{28}$$

$$W_n^f(z_n^*) \geq W_{n+1}^{d+}(z_n^*). \tag{29}$$

The internal stability condition in the steady state can be written out as

$$W_n^s(z_n^*) - W_{n-1}^{d-}(z_n^*) = \pi(q_n^{s*}) - \pi\left(q_{n-1}^{f*}(z_n^*)\right) \tag{30}$$

$$+\delta \frac{1}{n}\left[V_n^s(z_n^*) - V_n^s(z_{n-1}')\right] \geq 0.$$

16

The external stability condition in turn equals

$$W_{n+1}^{d+}\left(z_n^*\right) - W_n^f\left(z_n^*\right) = \pi\left(q_{n+1}^{s*}\left(z_n^*\right)\right) - \pi\left(q_n^{f*}\right) \tag{31}$$

$$+\delta\left[V_n^f\left(z_{n+1}'\right) - V_n^f\left(z_n^*\right)\right] \le 0.$$

The stable size of the agreement in the steady state will be the largest n^* that satisfies the internal and external stability conditions (30) and (31) at the steady state stock z_n^* corresponding to n.

3.3.2 Stability check 2: Infinite punishment to deviating signatories

We next consider stability when a signatory country leaving the agreement is not allowed to rejoin the coalition. If there are no profitable deviations where a signatory country i leaves the agreement in period t to remain a non-signatory thereafter, the membership size n is internally stable in the steady state corresponding to n. External stability still corresponds to there being no profitable deviations where a non-signatory country j joins the agreement in period t and exits the agreement in period $t+1$.

Suppose a signatory country i deviates from the agreement. Country i then emits as a non-signatory from the current period onwards, and remains a non-signatory up to infinity. The stock in the following period again obtains the value given by Equation (25). The continuation payoff to a signatory country deviating from the agreement then is [**notation - $W_{n-1}^{d-,\infty}$**

$$W_{n-1}^{d-,\infty}\left(z_n^*\right) = \pi\left(q_{n-1}^f\left(z_n^*\right)\right) - \gamma\frac{1}{2}\left(z_n^*\right)^2 + \delta V_{n-1}^f\left(z_{n-1}'\right).$$

Period $t+1$ stock and continuation payoff in the case of a one-time deviation by a non-signatory country j in period t are defined as in the previous section (Equations 26 and 27).

The internal stability condition in the steady state is

$$W_n^s\left(z_n^*\right) - W_{n-1}^{d-\infty}\left(z_n^*\right) = \pi\left(q_n^{s*}\right) - \pi\left(q_{n-1}^{f*}\left(z_n^*\right)\right) \tag{32}$$

$$+\delta\left[\frac{1}{n}V_n^s\left(z_n^*\right) - V_{n-1}^f\left(z_{n-1}'\right)\right] \ge 0.$$

The external stability condition is again given by equation (31).

4 Numerical application and results

Allowing for quadratic benefits and damages brings about need to use numerical analysis to solve the model. We consider a numerical example based on data pertaining to climate

change, from Nordhaus (1994), Tol (1995), Dellink et al. (2004) and Tol (2008). Given that this paper aims at a game-theoretic analysis of an IEA, we have made some simplifying assumptions regarding variety of climate change scenarios and ranges of uncertainty and focus on a limited number of parameter estimates in the baseline scenario. We consider a world of $N = 10$ identical regions. To obtain an estimate for parameter α that characterizes the marginal benefit from production, we utilize a result by Diamantoudi and Sartzetakis (2006) that a direct correspondence exist between emission and abatement models. That is, the slope parameter for marginal abatement costs equals the slope parameter for marginal benefits from emissions in the case of a quadratic payoff structure. Empirical literature has largely used the abatement model approach, rather then emission model, and we proceed by approximating a parameter that describes the slope of the marginal abatement function. We use available estimates of abatement functions provided by Dellink et al. (2004) and Nordhaus (1994), who use a cubic-quadratic functional form calibrated for twelve world regions. We select the regions with the highest and lowest emission paths according to their estimates, which are the USA and Brazil. Using the least squares method we obtain a range of values for parameter α, $\alpha \in [0.05, 19]$, where α has the dimension $\frac{10^9 \$}{(GtC)^2}$. To obtain an estimate for parameter γ that characterizes the slope of the marginal damage from the pollution stock, we use Tol's (1995) projections of global warming damage costs from 1990 till 2100, and Nordhaus and Sztorc (2013) data on atmospheric concentrations of CO_2 under alternative policies in DICE-2013R. This yields an approximation for parameter γ equal to $9.8 \cdot 10^{-6} \frac{\$}{tC}$. The discount factor is set equal to $\delta = 0.99$ and the carry-over coefficient to 0.92. Table 1 displays the parameter values used in the numerical analysis.

Table 1: Model parameters values

Parameter	value	units
Parameter of marginal benefit function α	$[0.05, 19]$	$\frac{10^9 \$}{GtC^2}$
Parameter of marginal damage function γ	$9.8 \cdot 10^{-6}$	$\frac{10^9 \$}{GtC^2}$
Carry over ρ	0.92	
Discount factor δ	0.99	
Initial stock of CO_2 in 2010	70	GtC
Number of countries (world regions)	10	

For all the parameter values in Table 1, there always exists a self-enforcing IEA in the steady state. We have explored several values for the damage parameter γ, ranging from

0.5 to 1.5 times the base line value $\gamma_b = 9.8 \cdot 10^{-6} \frac{\$}{tC}$. In the present numerical analysis, the results are obtained for $\alpha = 0.05$ and total number of players $N = 200$.

Table 2. Agreement for different values of γ, evaluated at $\alpha = 0.05$ and $N = 200$

γ	IS_1	IS_∞	ES_1	$n^{*,1}$	$n^{*,\infty}$
$4,9\ 10^{-6}$	2	2	23	2	2
$6,9\ 10^{-6}$	2	2	23	2	2
$7,8\ 10^{-6}$	2	2	23	2	2
$9,8\ 10^{-6}$	2	2	22	2	2
$11,8\ 10^{-6}$	2	2	22	2	2
$13,7\ 10^{-6}$	59	2	21	2	21
$14,7\ 10^{-6}$	48	2	21	2	21
$15,7\ 10^{-6}$	42	2	20	2	20
$18,6\ 10^{-6}$	31	2	19	2	19

The first column in Tab. [2] contains ranging values of γ, the second column of the table contains the largest number of signatories for which internal stability holds under one-shot deviation assumption, the third column contains the largest number of signatories for which internal stability holds under infinite punishment, the forth column contains the largest number of signatories for which external stability holds (one-shot deviation). The last two columns in Tab. [2] present equilibrium number IEA under one-shot assumption and infinite punishment assumption. Our results show that equilibrium size of the IEA membership strongly depends on whether a potential deviator may or may not be allowed to rejoin the agreement after the deviation: in case when a deviator can return to the agreement at the next stage following her deviation, the agreement at the steady state contains only two countries (similar to Rubio and Ulph (2007)), however if a deviator can not return, the agreement size can increase (if damage parameter γ is large enough).

5 Conclusions

This paper analyzes a dynamic model of international environmental cooperation, where the stock of pollution affects both signatory and non-signatory countries' emissions, and the stability of the IEA. We extend existing literature by allowing for quadratic benefits from emissions together with endogenous IEA membership. We apply the model to a numerical calibration reflecting the climate policy problem.

We find that that emission stock in fully cooperative case is below emission stock in non-cooperative, also discounted net benefits are lower in case of no cooperation than in case of cooperation. It is interesting that in cooperative case when α is low emission stock approaches steady state from above and when α is high it approaches steady state from below. Both cooperative and non-cooperative emission stocks approach steady state asymptotically. The following results differ from observations of similar cases in Rubio and Ulph (2007).

Firstly, it is interesting that in cooperative case when α is low emission stock approaches steady state from above and when α is high it approaches steady state from below. Both cooperative and non-cooperative emission stocks approach steady state asymptotically. Secondly, we can learn from our example that when α is low optimal emission path in cooperative case starts low and then goes up towards steady state, while optimal emission path in non-cooperative case starts high and then goes down towards steady state. When α is high optimal emission path in both cooperative and non-cooperative case start high and then goes down towards steady state. Thus while players behave in a similar way in the absence of cooperation, it case of full cooperation emission path and approach towards steady state of the stock largely depend on parameter α of benefit function.

Further, we analyze effect of pollutant stock dynamics on IEA membership assuming that at any time period each country makes a decision whether to participate in an IEA or not, and depending on the outcome on these decisions, each country adjusts its emission policy according to the feedback Nash equilibrium. Our objective is to find equilibrium size of an IEA, which withstands possible one-shot deviations. The model is applied to a numerical calibration reflecting the climate policy problem. We show that optimal emission policy is determined as a function of the current pollution level and equilibrium size of the IEA membership strongly depends on whether a potential deviator may or may not be allowed to rejoin the agreement after the deviation: in case when a deviator can return to the agreement at the next stage following her deviation, the agreement at the steady state contains only two countries, however if a deviator can not return, the agreement size can be very large.

References

d'Aspremont, C., A. Jacquemin, J. Jaskold-Gabszewicz and J. Weymark, 1983. On the Stability of Collusive Price Leadership, Canadian Journal of Economics, 16: 17-25.

Barrett, S., 1997. Heterogeneous International Environmental Agreements, in Carlo Carraro (Ed), International Environmental Negotiations: Strategic Policy Issues (Edward Elgar, Cheltenham, UK, 1997a).

Barrett, S. Environment and Statecraft (Oxford University Press, Oxford, 2003).

Battaglini, M. and Harstad, B. (2015), Participation and Duration of Environmental Agreements, forthcoming, Journal of Political Economy.

Breton, M., Sbragia, L., and Zaccour, G., 2010. Dynamic models for international environmental agreements. Environmental and Resource Economic, 45(1): 25-48.

Carraro, C. and Siniscalco, D. 1993. Strategies for the international protection of the environment. Journal of Public Economics, 52(3): 309-328.

Nordhaus, William, and Paul Sztorc. DICE 2013R: Introduction and user's manual, 2013., available at http://www.econ.yale.edu/~nordhaus/homepage/documents/

DICE_Manual_103113r2.pdf.

Rubio, S., and Casino, B., 2005. Self-enforcing international environmental agreements with a stock pollutant. Spanish Economic Review, 7(2): 89-109.

Rubio, S., and Ulph, A. (2007). An infinite-horizon model of dynamic membership of international environmental agreements. Journal of Environmental Economics and Management, 54(3): 296-310.

6   Appendix

A signatory's payoff from deviating (with return to cooperation in period $t + 1$) is

$$
J\left(z_t\right) = \pi\left[q_n^s\left(z_t\right)\right] - \pi\left[q_{n-1}^f\left(z_t\right)\right]
$$
$$
+\delta\frac{1}{n}\left[V_n^s\left(z_t\right) - V_n^s\left(z_{t+1,n-1}\right)\right]
$$

$$
= \pi\left[q_n^s\left(z_t\right)\right] - \pi\left[q_{n-1}^f\left(z_t\right)\right]
$$
$$
+\delta\frac{1}{n}V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]
$$
$$
-\delta\frac{1}{n}V_n^s\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]
$$

Differentiating with respect to z_t yieldsη

$$\frac{\partial J\left(z_t\right)}{\partial z_t} = \frac{\partial \pi\left[q_n^s\left(z_t\right)\right]}{\partial q_n^s} \frac{dq_n^s\left(z_t\right)}{dz_t} - \frac{\partial \pi\left[q_{n-1}^f\left(z_t\right)\right]}{\partial q_{n-1}^f\left(z_t\right)} \frac{dq_{n-1}^f\left(z_t\right)}{dz_t} \tag{33}$$

$$+\delta\rho\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial z_{t+1}}$$

$$+\delta\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial z_{t+1}}\frac{dz_{t+1}}{dq_n^s}\frac{dq_n^s\left(z_t\right)}{dz_t}$$

$$+\delta\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial z_{t+1}}\frac{dz_{t+1}}{dq_n^f}\frac{dq_n^f\left(z_t\right)}{dz_t}$$

$$-\delta\rho\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]}{\partial z_{t+1}}$$

$$-\delta\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]}{\partial z_{t+1}}\frac{dz_{t+1}}{dq_{n-1}^s}\frac{dq_{n-1}^s\left(z_t\right)}{dz_t}$$

$$-\delta\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]}{\partial z_{t+1}}\frac{dz_{t+1}}{dq_{n-1}^f}\frac{dq_{n-1}^f\left(z_t\right)}{dz_t}$$

$$\frac{\partial J\left(z_t\right)}{\partial z_t} = \frac{\partial \pi\left[q_n^s\left(z_t\right)\right]}{\partial q_n^s}\eta_{s,n} - \frac{\partial \pi\left[q_{n-1}^f\left(z_t\right)\right]}{\partial q_{n-1}^f\left(z_t\right)}\eta_{f,n-1} \tag{34}$$

$$+\delta\rho\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial z_{t+1}}$$

$$+\delta\frac{1}{n}n\frac{\partial V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial z_{t+1}}\eta_{s,n}$$

$$+\delta\frac{1}{n}(N-n)\frac{\partial V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial z_{t+1}}\eta_{f,n}$$

$$-\delta\rho\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]}{\partial z_{t+1}}$$

$$-\delta\frac{1}{n}(n-1)\frac{\partial V_n^s\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]}{\partial z_{t+1}}\eta_{s,n-1}$$

$$-\delta\frac{1}{n}(N-n+1)\frac{\partial V_n^s\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]}{\partial z_{t+1}}\eta_{f,n-1}$$

$$\frac{\partial^2 J\left(z_t\right)}{\partial z_t^2} = \left(\eta_{s,n}\right)^2 \frac{\partial^2 \pi\left[q_n^s\left(z_t\right)\right]}{\partial^2 q_n^s} - \left(\eta_{f,n-1}\right)^2 \frac{\partial \pi^2\left[q_{n-1}^f\left(z_t\right)\right]}{\partial^2 q_{n-1}^f\left(z_t\right)} \tag{35}$$

$$+\delta\rho^2\frac{1}{n}\frac{\partial^2 V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial^2 z_{t+1}}$$

$$+\delta\frac{1}{n}n\eta_{s,n}\frac{\partial^2 V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial^2 z_{t+1}}\frac{dz_{t+1}}{dq_n^s}\frac{dq_n^s\left(z_t\right)}{dz_t} \tag{36}$$

$$+\delta\frac{1}{n}(N-n)\eta_{f,n}\frac{\partial^2 V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial^2 z_{t+1}}\frac{dz_{t+1}}{dq_n^f}\frac{dq_n^f\left(z_t\right)}{dz_t} \tag{37}$$

$$-\delta\rho^2\frac{1}{n}\frac{\partial^2 V_n^s\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]}{\partial^2 z_{t+1}} \tag{38}$$

$$-\delta\frac{1}{n}(n-1)\eta_{s,n-1}\frac{\partial^2 V_n^s\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]}{\partial^2 z_{t+1}}\frac{dz_{t+1}}{dq_{n-1}^s}\frac{dq_{n-1}^s\left(z_t\right)}{dz_t}$$

$$\tag{39}$$

$$-\delta\frac{1}{n}(N-n+1)\eta_{f,n-1}\frac{\partial V_n^s\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]}{\partial z_{t+1}}\frac{dz_{t+1}}{dq_{n-1}^f}\frac{dq_{n-1}^f\left(z_t\right)}{dz_t}$$

$$\tag{40}$$

$$\frac{\partial^2 J\left(z_t\right)}{\partial z_t^2} = -\left(\eta_{s,n}\right)^2\alpha + \left(\eta_{f,n-1}\right)^2\alpha$$

$$+\delta\rho^2\frac{1}{n}A_n^s$$

$$+\delta\frac{1}{n}n^2\left(\eta_{s,n}\right)^2 A_n^s + \delta\frac{1}{n}(N-n)^2\left(\eta_{f,n}^2\right)^2 A_n^s$$

$$-\delta\rho^2\frac{1}{n}A_n^s$$

$$-\delta\frac{1}{n}(n-1)^2\left(\eta_{s,n-1}\right)^2 A_n^s - \delta\frac{1}{n}(N-n+1)^2\left(\eta_{f,n-1}\right)^2 A_n^s$$

$$\frac{\partial^2 J\left(z_t\right)}{\partial z_t^2} = \alpha\left[\left(\eta_{f,n-1}\right)^2 - \left(\eta_{s,n}\right)^2\right]$$

$$+A_n^s\left[\delta n\left(\eta_{s,n}\right)^2 + \delta\frac{1}{n}(N-n)^2\left(\eta_{f,n}^2\right)^2\right.$$

$$\left.-\delta\frac{1}{n}(n-1)^2\left(\eta_{s,n-1}\right)^2 - \delta\frac{1}{n}(N-n+1)^2\left(\eta_{f,n-1}\right)^2\right]$$

$$\frac{\partial^2 J\left(z_t\right)}{\partial z_t^2} = \alpha \left[\left(\eta_{f,n-1}\right)^2 - \left(\eta_{s,n}\right)^2\right]$$

$$+A_n^s \delta \frac{1}{n}\left[n^2\left(\eta_{s,n}\right)^2 - (n-1)^2\left(\eta_{s,n-1}\right)^2\right.$$

$$\left.+(N-n)^2\left(\eta_{f,n}\right)^2 - (N-n+1)^2\left(\eta_{f,n-1}\right)^2\right]$$

OK UP TO HERE

CHeck this -could perhaps get rid off the first two lines using the envelope theorem. Could there be some argument for why $A_n^s >$ or $< A_n^f$?

$$\frac{\partial \Phi\left(z_t\right)}{\partial z_t} = \left\{\frac{\partial \pi\left[q_n^s\left(z_t\right)\right]}{\partial q_n^s} + \delta n \frac{\partial V_n^s\left[\rho z_t + n q_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial z_{t+1}}\right\}\frac{dq_n^s\left(z_t\right)}{dz_t} \qquad (44)$$

$$-\left\{\frac{\partial \pi\left[q_{n-1}^f\left(z_t\right)\right]}{\partial q_{n-1}^f\left(z_t\right)} + \delta \frac{\partial V_n^s\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]}{\partial z_{t+1}}\right\}\frac{dq_{n-1}^f\left(z_t\right)}{dz_t}$$

$$\tag{45}$$

$$+\delta\rho\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + n q_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial z_{t+1}} \qquad (46)$$

$$-\delta(n-1)\frac{\partial V_n^s\left[\rho z_t + n q_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial z_{t+1}}\frac{dq_n^s\left(z_t\right)}{dz_t}$$

$$+\delta\frac{N-n}{n}\frac{\partial V_n^s\left[\rho z_t + n q_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial z_{t+1}}\frac{dq_n^f\left(z_t\right)}{dz_t}$$

$$-\delta\rho\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]}{\partial z_{t+1}}$$

$$-\delta\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]}{\partial z_{t+1}}\frac{dz_{t+1}}{dq_{n-1}^s}\frac{dq_{n-1}^s\left(z_t\right)}{dz_t}$$

$$-\delta\frac{(N-2n+1)}{n}\frac{\partial V_n^s\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]}{\partial z_{t+1}}\frac{dq_{n-1}^f\left(z_t\right)}{dz_t} \qquad (47)$$

$$\tag{48}$$

$$\frac{\partial \Phi\left(z_t\right)}{\partial z_t} = \frac{\partial \pi\left[q_n^s\left(z_t\right)\right]}{\partial q_n^s} \frac{dq_n^s\left(z_t\right)}{dz_t} - \frac{\partial \pi\left[q_{n-1}^f\left(z_t\right)\right]}{\partial q_{n-1}^f\left(z_t\right)} \frac{dq_{n-1}^f\left(z_t\right)}{dz_t}$$

$$+\delta\rho\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial z_{t+1}}$$

$$+\delta\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial z_{t+1}}\frac{dz_{t+1}}{dq_n^s}\frac{dq_n^s\left(z_t\right)}{dz_t} \qquad (49)$$

A signatory's payoff from deviating at ns=2 (with return to cooperation in period $t+1$) is

$$Jns2\left(z_t\right) = \pi\left[q_n^s\left(z_t\right)\right] - \pi\left[q_{nc}\left(z_t\right)\right]$$

$$+\delta\frac{1}{n}\left[V_n^s\left(z_t\right) - V_n^s\left(z_{t+1,n-1}\right)\right]$$

with

$$z_{t+1,n-1} = \rho z_t + Nq_{nc}\left(z_t\right)$$

$$Jns2\left(z_t\right) = \pi\left[q_n^s\left(z_t\right)\right] - \pi\left[q_{nc}\left(z_t\right)\right]$$

$$+\delta\frac{1}{n}V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]$$

$$-\delta\frac{1}{n}V_n^s\left[\rho z_t + Nq_{nc}\left(z_t\right)\right]$$

Differentiating with respect to z_t yieldsη

$$\frac{\partial ns2J\left(z_t\right)}{\partial z_t} = \frac{\partial \pi\left[q_n^s\left(z_t\right)\right]}{\partial q_n^s} \frac{dq_n^s\left(z_t\right)}{dz_t} - \frac{\partial \pi\left[q_{nc}\left(z_t\right)\right]}{\partial q_{nc}\left(z_t\right)} \frac{dq_{nc}\left(z_t\right)}{dz_t}$$

$$+\delta\rho\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial z_{t+1}}$$

$$+\delta\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial z_{t+1}}\frac{dz_{t+1}}{dq_n^s}\frac{dq_n^s\left(z_t\right)}{dz_t}$$

$$+\delta\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial z_{t+1}}\frac{dz_{t+1}}{dq_n^f}\frac{dq_n^f\left(z_t\right)}{dz_t}$$

$$-\delta\rho\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]}{\partial z_{t+1}}$$

$$-\delta\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]}{\partial z_{t+1}}\frac{dz_{t+1}}{dq_{nc}}\frac{dq_{nc}\left(z_t\right)}{dz_t} \qquad (50)$$

25

$$\frac{\partial Jns2\,(z_t)}{\partial z_t} = \frac{\partial \pi\,[q_n^s\,(z_t)]}{\partial q_n^s}\eta_{s,n} - \frac{\partial \pi\,[q_{nc}\,(z_t)]}{\partial q_{nc}\,(z_t)}\eta_{nc} \tag{51}$$

$$+\delta\rho\frac{1}{n}\frac{\partial V_n^s\,[\rho z_t + n q_n^s\,(z_t) + (N-n)q_n^f\,(z_t)]}{\partial z_{t+1}}$$

$$+\delta\frac{1}{n}n\frac{\partial V_n^s\,[\rho z_t + n q_n^s\,(z_t) + (N-n)q_n^f\,(z_t)]}{\partial z_{t+1}}\eta_{s,n}$$

$$+\delta\frac{1}{n}(N-n)\frac{\partial V_n^s\,[\rho z_t + n q_n^s\,(z_t) + (N-n)q_n^f\,(z_t)]}{\partial z_{t+1}}\eta_{f,n}$$

$$-\delta\rho\frac{1}{n}\frac{\partial V_n^s\,[\rho z_t + N q_{nc}\,(z_t)]}{\partial z_{t+1}}$$

$$-\delta\frac{1}{n}N\frac{\partial V_n^s\,[\rho z_t + N q_{nc}\,(z_t)]}{\partial z_{t+1}}\eta_{nc}$$

$$\frac{\partial^2 Jns2\,(z_t)}{\partial z_t^2} = (\eta_{s,n})^2\,\frac{\partial^2 \pi\,[q_n^s\,(z_t)]}{\partial^2 q_n^s} - (\eta_{nc})^2\,\frac{\partial^2 \pi\,[q_{nc}\,(z_t)]}{\partial^2 q_{nc}\,(z_t)} \tag{52}$$

$$+\delta\rho^2\frac{1}{n}\frac{\partial^2 V_n^s\,[\rho z_t + n q_n^s\,(z_t) + (N-n)q_n^f\,(z_t)]}{\partial^2 z_{t+1}}$$

$$+\delta\frac{1}{n}n\eta_{s,n}\frac{\partial^2 V_n^s\,[\rho z_t + n q_n^s\,(z_t) + (N-n)q_n^f\,(z_t)]}{\partial^2 z_{t+1}}\frac{dz_{t+1}}{dq_n^s}\frac{dq_n^s\,(z_t)}{dz_t} \tag{53}$$

$$+\delta\frac{1}{n}(N-n)\eta_{f,n}\frac{\partial^2 V_n^s\,[\rho z_t + n q_n^s\,(z_t) + (N-n)q_n^f\,(z_t)]}{\partial^2 z_{t+1}}\frac{dz_{t+1}}{dq_n^f}\frac{dq_n^f\,(z_t)}{dz_t} \tag{54}$$

$$-\delta\rho^2\frac{1}{n}\frac{\partial^2 V_n^s\,[\rho z_t + (n-1)q_{n-1}^s\,(z_t) + (N-n+1)q_{n-1}^f\,(z_t)]}{\partial^2 z_{t+1}} \tag{55}$$

$$-\delta\frac{1}{n}N\eta_{nc}\frac{\partial^2 V_n^s\,[\rho z_t + (n-1)q_{n-1}^s\,(z_t) + (N-n+1)q_{n-1}^f\,(z_t)]}{\partial^2 z_{t+1}}\frac{dz_{t+1}}{dq_{nc}}\frac{dq_{nc}\,(z_t)}{dz_t} \tag{56}$$

$$\frac{\partial^2 Jns2\,(z_t)}{\partial z_t^2} = -(\eta_{s,n})^2\,\alpha + (\eta_{nc})^2\,\alpha$$

$$+\delta\rho^2\frac{1}{n}A_n^s$$

$$+\delta\frac{1}{n}n^2\,(\eta_{s,n})^2\,A_n^s + \delta\frac{1}{n}(N-n)^2\,(\eta_{f,n})^2\,A_n^s$$

$$-\delta\rho^2\frac{1}{n}A_n^s - \delta\frac{1}{n}N^2\,(\eta_{nc})^2\,A_n^s$$

$$\frac{\partial^2 ns2J(z_t)}{\partial z_t^2} = \alpha\left[(\eta_{nc})^2 - (\eta_{s,n})^2\right]$$

$$+A_n^s\left[\delta n(\eta_{s,n})^2 + \delta\frac{1}{n}(N-n)^2\left(\eta_{f,n}^2\right)^2\right.$$

$$\left.-\delta\frac{1}{n}N^2(\eta_{nc})^2\right]$$

$$\frac{\partial^2 Jns2(z_t)}{\partial z_t^2} = \alpha\left[(\eta_{nc})^2 - (\eta_{s,n})^2\right]$$

$$+A_n^s\delta\frac{1}{n}\left[n^2(\eta_{s,n})^2 + (N-n)^2(\eta_{f,n})^2\right.$$

$$\left.-N^2(\eta_{nc})^2\right]$$

OK UP TO HERE

CHeck this -could perhaps get rid off the first two lines using the envelope theorem. Could there be some argument for why $A_n^s >$ or $< A_n^f$?

$$\frac{\partial\Phi(z_t)}{\partial z_t} = \left\{\frac{\partial\pi\left[q_n^s(z_t)\right]}{\partial q_n^s} + \delta n\frac{\partial V_n^s\left[\rho z_t + nq_n^s(z_t) + (N-n)q_n^f(z_t)\right]}{\partial z_{t+1}}\right\}\frac{dq_n^s(z_t)}{dz_t} \quad (60)$$

$$-\left\{\frac{\partial\pi\left[q_{n-1}^f(z_t)\right]}{\partial q_{n-1}^f(z_t)} + \delta\frac{\partial V_n^s\left[\rho z_t + (n-1)q_{n-1}^s(z_t) + (N-n+1)q_{n-1}^f(z_t)\right]}{\partial z_{t+1}}\right\}\frac{dq_{n-1}^f(z_t)}{dz_t}$$

$$(61)$$

$$+\delta\rho\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + nq_n^s(z_t) + (N-n)q_n^f(z_t)\right]}{\partial z_{t+1}} \quad (62)$$

$$-\delta(n-1)\frac{\partial V_n^s\left[\rho z_t + nq_n^s(z_t) + (N-n)q_n^f(z_t)\right]}{\partial z_{t+1}}\frac{dq_n^s(z_t)}{dz_t}$$

$$+\delta\frac{N-n}{n}\frac{\partial V_n^s\left[\rho z_t + nq_n^s(z_t) + (N-n)q_n^f(z_t)\right]}{\partial z_{t+1}}\frac{dq_n^f(z_t)}{dz_t}$$

$$-\delta\rho\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + (n-1)q_{n-1}^s(z_t) + (N-n+1)q_{n-1}^f(z_t)\right]}{\partial z_{t+1}}$$

$$-\delta\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + (n-1)q_{n-1}^s(z_t) + (N-n+1)q_{n-1}^f(z_t)\right]}{\partial z_{t+1}}\frac{dz_{t+1}}{dq_{n-1}^s}\frac{dq_{n-1}^s(z_t)}{dz_t}$$

$$-\delta\frac{(N-2n+1)}{n}\frac{\partial V_n^s\left[\rho z_t + (n-1)q_{n-1}^s(z_t) + (N-n+1)q_{n-1}^f(z_t)\right]}{\partial z_{t+1}}\frac{dq_{n-1}^f(z_t)}{dz_t} \quad (63)$$

$$(64)$$

$$\frac{\partial \Phi\left(z_t\right)}{\partial z_t} = \frac{\partial \pi\left[q_n^s\left(z_t\right)\right]}{\partial q_n^s} \frac{dq_n^s\left(z_t\right)}{dz_t} - \frac{\partial \pi\left[q_{n-1}^f\left(z_t\right)\right]}{\partial q_{n-1}^f\left(z_t\right)} \frac{dq_{n-1}^f\left(z_t\right)}{dz_t} \tag{65}$$

$$+ \delta\rho\frac{1}{n} \frac{\partial V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial z_{t+1}}$$

$$+ \delta\frac{1}{n} \frac{\partial V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial z_{t+1}} \frac{dz_{t+1}}{dq_n^s} \frac{dq_n^s\left(z_t\right)}{dz_t}$$

A signatory's payoff from deviating with "infinite punishment" is

$$J_{IP}\left(z_t\right) = \pi\left[q_n^s\left(z_t\right)\right] - \pi\left[q_{n-1}^f\left(z_t\right)\right]$$

$$+ \delta\left[\frac{1}{n}V_n^s\left(z_t\right) - V_{n-1}^f\left(z_{t+1,n-1}\right)\right]$$

$$= \pi\left[q_n^s\left(z_t\right)\right] - \pi\left[q_{n-1}^f\left(z_t\right)\right]$$

$$+ \delta\frac{1}{n}V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]$$

$$- \delta V_n^f\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]$$

Differentiating with respect to z_t yieldsη

$$\frac{\partial J_{IP}\left(z_t\right)}{\partial z_t} = \frac{\partial \pi\left[q_n^s\left(z_t\right)\right]}{\partial q_n^s} \frac{dq_n^s\left(z_t\right)}{dz_t} - \frac{\partial \pi\left[q_{n-1}^f\left(z_t\right)\right]}{\partial q_{n-1}^f\left(z_t\right)} \frac{dq_{n-1}^f\left(z_t\right)}{dz_t}$$

$$+ \delta\rho\frac{1}{n} \frac{\partial V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial z_{t+1}}$$

$$+ \delta\frac{1}{n} \frac{\partial V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial z_{t+1}} \frac{dz_{t+1}}{dq_n^s} \frac{dq_n^s\left(z_t\right)}{dz_t}$$

$$+ \delta\frac{1}{n} \frac{\partial V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial z_{t+1}} \frac{dz_{t+1}}{dq_n^f} \frac{dq_n^f\left(z_t\right)}{dz_t}$$

$$- \delta\rho\frac{\partial V_{n-1}^f\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]}{\partial z_{t+1}}$$

$$- \delta\frac{\partial V_{n-1}^f\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]}{\partial z_{t+1}} \frac{dz_{t+1}}{dq_{n-1}^s} \frac{dq_{n-1}^s\left(z_t\right)}{dz_t}$$

$$- \delta\frac{\partial V_{n-1}^f\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]}{\partial z_{t+1}} \frac{dz_{t+1}}{dq_{n-1}^f} \frac{dq_{n-1}^f\left(z_t\right)}{dz_t}$$

$$\frac{\partial J_{IP}(z_t)}{\partial z_t} = \frac{\partial \pi \left[q_n^s (z_t) \right]}{\partial q_n^s} \eta_{s,n} - \frac{\partial \pi \left[q_{n-1}^f (z_t) \right]}{\partial q_{n-1}^f (z_t)} \eta_{f,n-1} \tag{66}$$

$$+ \delta \rho \frac{1}{n} \frac{\partial V_n^s \left[\rho z_t + n q_n^s (z_t) + (N-n) q_n^f (z_t) \right]}{\partial z_{t+1}}$$

$$+ \delta \frac{1}{n} n \frac{\partial V_n^s \left[\rho z_t + n q_n^s (z_t) + (N-n) q_n^f (z_t) \right]}{\partial z_{t+1}} \eta_{s,n}$$

$$+ \delta \frac{1}{n} (N-n) \frac{\partial V_n^s \left[\rho z_t + n q_n^s (z_t) + (N-n) q_n^f (z_t) \right]}{\partial z_{t+1}} \eta_{f,n}$$

$$- \delta \rho \frac{\partial V_{n-1}^f \left[\rho z_t + (n-1) q_{n-1}^s (z_t) + (N-n+1) q_{n-1}^f (z_t) \right]}{\partial z_{t+1}}$$

$$- \delta (n-1) \frac{\partial V_{n-1}^f \left[\rho z_t + (n-1) q_{n-1}^s (z_t) + (N-n+1) q_{n-1}^f (z_t) \right]}{\partial z_{t+1}} \eta_{s,n-1}$$

$$- \delta (N-n+1) \frac{\partial V_{n-1}^f \left[\rho z_t + (n-1) q_{n-1}^s (z_t) + (N-n+1) q_{n-1}^f (z_t) \right]}{\partial z_{t+1}} \eta_{f,n-1}$$

$$\frac{\partial^2 J_{IP}(z_t)}{\partial z_t^2} = (\eta_{s,n})^2 \frac{\partial^2 \pi \left[q_n^s (z_t) \right]}{\partial^2 q_n^s} - (\eta_{f,n-1})^2 \frac{\partial \pi^2 \left[q_{n-1}^f (z_t) \right]}{\partial^2 q_{n-1}^f (z_t)} \tag{67}$$

$$+ \delta \rho^2 \frac{1}{n} \frac{\partial^2 V_n^s \left[\rho z_t + n q_n^s (z_t) + (N-n) q_n^f (z_t) \right]}{\partial^2 z_{t+1}}$$

$$+ \delta \frac{1}{n} n \eta_{s,n} \frac{\partial^2 V_n^s \left[\rho z_t + n q_n^s (z_t) + (N-n) q_n^f (z_t) \right]}{\partial^2 z_{t+1}} \frac{dz_{t+1}}{dq_n^s} \frac{dq_n^s (z_t)}{dz_t} \tag{68}$$

$$+ \delta \frac{1}{n} (N-n) \eta_{f,n} \frac{\partial^2 V_n^s \left[\rho z_t + n q_n^s (z_t) + (N-n) q_n^f (z_t) \right]}{\partial^2 z_{t+1}} \frac{dz_{t+1}}{dq_n^f} \frac{dq_n^f (z_t)}{dz_t} \tag{69}$$

$$- \delta \rho^2 \frac{\partial^2 V_{n-1}^f \left[\rho z_t + (n-1) q_{n-1}^s (z_t) + (N-n+1) q_{n-1}^f (z_t) \right]}{\partial^2 z_{t+1}} \tag{70}$$

$$- \delta (n-1) \eta_{s,n-1} \frac{\partial^2 V_{n-1}^f \left[\rho z_t + (n-1) q_{n-1}^s (z_t) + (N-n+1) q_{n-1}^f (z_t) \right]}{\partial^2 z_{t+1}} \frac{dz_{t+1}}{dq_{n-1}^s} \frac{dq_{n-1}^s (z_t)}{dz_t}$$

$$\tag{71}$$

$$- \delta (N-n+1) \eta_{f,n-1} \frac{\partial V_{n-1}^f \left[\rho z_t + (n-1) q_{n-1}^s (z_t) + (N-n+1) q_{n-1}^f (z_t) \right]}{\partial z_{t+1}} \frac{dz_{t+1}}{dq_{n-1}^f} \frac{dq_{n-1}^f (z_t)}{dz_t}$$

$$\tag{72}$$

$$\frac{\partial^2 J_{IP}(z_t)}{\partial z_t^2} = -\left(\eta_{s,n}\right)^2 \alpha + \left(\eta_{f,n-1}\right)^2 \alpha$$

$$+\delta \frac{1}{n} A_n^s \rho^2$$

$$+\delta \frac{1}{n} A_n^s n^2 \left(\eta_{s,n}\right)^2 + \delta \frac{1}{n} A_n^s (N-n)^2 \left(\eta_{f,n}^2\right)^2$$

$$-\delta \rho^2 A_{n-1}^f$$

$$-\delta(n-1)^2 \left(\eta_{s,n-1}\right)^2 A_{n-1}^f - \delta(N-n+1)^2 \left(\eta_{f,n-1}\right)^2 A_{n-1}^f$$

$$\frac{\partial^2 J_{IP}(z_t)}{\partial z_t^2} = -\left(\eta_{s,n}\right)^2 \alpha + \left(\eta_{f,n-1}\right)^2 \alpha$$

$$+\delta \frac{1}{n} A_n^s \left[\rho^2 + n^2 \left(\eta_{s,n}\right)^2 + (N-n)^2 \left(\eta_{f,n}^2\right)^2\right] \tag{74a}$$

$$-\delta A_{n-1}^f \left[\rho^2 - (n-1)^2 \left(\eta_{s,n-1}\right)^2 - (N-n+1)^2 \left(\eta_{f,n-1}\right)^2\right]$$

OK UP TO HERE

CHeck this -could perhaps get rid off the first two lines using the envelope theorem. Could there be some argument for why $A_n^s >$ or $< A_n^f$?

$$\frac{\partial \Phi\left(z_t\right)}{\partial z_t} = \left\{ \frac{\partial \pi\left[q_n^s\left(z_t\right)\right]}{\partial q_n^s} + \delta n \frac{\partial V_n^s\left[\rho z_t + n q_n^s\left(z_t\right) + (N-n) q_n^f\left(z_t\right)\right]}{\partial z_{t+1}} \right\} \frac{dq_n^s\left(z_t\right)}{dz_t} \tag{75}$$

$$-\left\{ \frac{\partial \pi\left[q_{n-1}^f\left(z_t\right)\right]}{\partial q_{n-1}^f\left(z_t\right)} + \delta \frac{\partial V_n^s\left[\rho z_t + (n-1) q_{n-1}^s\left(z_t\right) + (N-n+1) q_{n-1}^f\left(z_t\right)\right]}{\partial z_{t+1}} \right\} \frac{dq_{n-1}^f\left(z_t\right)}{dz_t} \tag{76}$$

$$+\delta\rho \frac{1}{n} \frac{\partial V_n^s\left[\rho z_t + n q_n^s\left(z_t\right) + (N-n) q_n^f\left(z_t\right)\right]}{\partial z_{t+1}} \tag{77}$$

$$-\delta(n-1) \frac{\partial V_n^s\left[\rho z_t + n q_n^s\left(z_t\right) + (N-n) q_n^f\left(z_t\right)\right]}{\partial z_{t+1}} \frac{dq_n^s\left(z_t\right)}{dz_t}$$

$$+\delta \frac{N-n}{n} \frac{\partial V_n^s\left[\rho z_t + n q_n^s\left(z_t\right) + (N-n) q_n^f\left(z_t\right)\right]}{\partial z_{t+1}} \frac{dq_n^f\left(z_t\right)}{dz_t}$$

$$-\delta\rho \frac{1}{n} \frac{\partial V_n^s\left[\rho z_t + (n-1) q_{n-1}^s\left(z_t\right) + (N-n+1) q_{n-1}^f\left(z_t\right)\right]}{\partial z_{t+1}}$$

$$-\delta \frac{1}{n} \frac{\partial V_n^s\left[\rho z_t + (n-1) q_{n-1}^s\left(z_t\right) + (N-n+1) q_{n-1}^f\left(z_t\right)\right]}{\partial z_{t+1}} \frac{dz_{t+1}}{dq_{n-1}^s} \frac{dq_{n-1}^s\left(z_t\right)}{dz_t}$$

$$-\delta \frac{(N-2n+1)}{n} \frac{\partial V_n^s\left[\rho z_t + (n-1) q_{n-1}^s\left(z_t\right) + (N-n+1) q_{n-1}^f\left(z_t\right)\right]}{\partial z_{t+1}} \frac{dq_{n-1}^f\left(z_t\right)}{dz_t} \tag{78}$$

$$\tag{79}$$

$$\frac{\partial \Phi\left(z_t\right)}{\partial z_t} = \frac{\partial \pi\left[q_n^s\left(z_t\right)\right]}{\partial q_n^s} \frac{dq_n^s\left(z_t\right)}{dz_t} - \frac{\partial \pi\left[q_{n-1}^f\left(z_t\right)\right]}{\partial q_{n-1}^f\left(z_t\right)} \frac{dq_{n-1}^f\left(z_t\right)}{dz_t} \tag{80}$$

$$+\delta\rho \frac{1}{n} \frac{\partial V_n^s\left[\rho z_t + n q_n^s\left(z_t\right) + (N-n) q_n^f\left(z_t\right)\right]}{\partial z_{t+1}}$$

$$+\delta \frac{1}{n} \frac{\partial V_n^s\left[\rho z_t + n q_n^s\left(z_t\right) + (N-n) q_n^f\left(z_t\right)\right]}{\partial z_{t+1}} \frac{dz_{t+1}}{dq_n^s} \frac{dq_n^s\left(z_t\right)}{dz_t}$$

A signatory's payoff from deviating at ns=2 (with return to cooperation in period $t+1$) is

$$Jns2\left(z_t\right) = \pi\left[q_n^s\left(z_t\right)\right] - \pi\left[q_{nc}\left(z_t\right)\right]$$

$$+\delta \frac{1}{n}\left[V_n^s\left(z_t\right) - V_n^s\left(z_{t+1,n-1}\right)\right]$$

with

$$z_{t+1,n-1} = \rho z_t + N q_{nc}(z_t)$$

$$Jns2(z_t) = \pi[q_n^s(z_t)] - \pi[q_{nc}(z_t)]$$
$$+\delta \frac{1}{n} V_n^s[\rho z_t + n q_n^s(z_t) + (N-n)q_n^f(z_t)]$$
$$-\delta \frac{1}{n} V_n^s[\rho z_t + N q_{nc}(z_t)]$$

Differentiating with respect to z_t yieldsη

$$\frac{\partial ns2J(z_t)}{\partial z_t} = \frac{\partial \pi[q_n^s(z_t)]}{\partial q_n^s} \frac{dq_n^s(z_t)}{dz_t} - \frac{\partial \pi[q_{nc}(z_t)]}{\partial q_{nc}(z_t)} \frac{dq_{nc}(z_t)}{dz_t} \tag{81}$$
$$+\delta \rho \frac{1}{n} \frac{\partial V_n^s[\rho z_t + n q_n^s(z_t) + (N-n)q_n^f(z_t)]}{\partial z_{t+1}}$$
$$+\delta \frac{1}{n} \frac{\partial V_n^s[\rho z_t + n q_n^s(z_t) + (N-n)q_n^f(z_t)]}{\partial z_{t+1}} \frac{dz_{t+1}}{dq_n^s} \frac{dq_n^s(z_t)}{dz_t}$$
$$+\delta \frac{1}{n} \frac{\partial V_n^s[\rho z_t + n q_n^s(z_t) + (N-n)q_n^f(z_t)]}{\partial z_{t+1}} \frac{dz_{t+1}}{dq_n^f} \frac{dq_n^f(z_t)}{dz_t}$$
$$-\delta \rho \frac{1}{n} \frac{\partial V_n^s[\rho z_t + (n-1)q_{n-1}^s(z_t) + (N-n+1)q_{n-1}^f(z_t)]}{\partial z_{t+1}}$$
$$-\delta \frac{1}{n} \frac{\partial V_n^s[\rho z_t + (n-1)q_{n-1}^s(z_t) + (N-n+1)q_{n-1}^f(z_t)]}{\partial z_{t+1}} \frac{dz_{t+1}}{dq_{nc}} \frac{dq_{nc}(z_t)}{dz_t}$$

$$\frac{\partial Jns2(z_t)}{\partial z_t} = \frac{\partial \pi[q_n^s(z_t)]}{\partial q_n^s} \eta_{s,n} - \frac{\partial \pi[q_{nc}(z_t)]}{\partial q_{nc}(z_t)} \eta_{nc} \tag{82}$$
$$+\delta \rho \frac{1}{n} \frac{\partial V_n^s[\rho z_t + n q_n^s(z_t) + (N-n)q_n^f(z_t)]}{\partial z_{t+1}}$$
$$+\delta \frac{1}{n} n \frac{\partial V_n^s[\rho z_t + n q_n^s(z_t) + (N-n)q_n^f(z_t)]}{\partial z_{t+1}} \eta_{s,n}$$
$$+\delta \frac{1}{n}(N-n) \frac{\partial V_n^s[\rho z_t + n q_n^s(z_t) + (N-n)q_n^f(z_t)]}{\partial z_{t+1}} \eta_{f,n}$$
$$-\delta \rho \frac{1}{n} \frac{\partial V_n^s[\rho z_t + N q_{nc}(z_t)]}{\partial z_{t+1}}$$
$$-\delta \frac{1}{n} N \frac{\partial V_n^s[\rho z_t + N q_{nc}(z_t)]}{\partial z_{t+1}} \eta_{nc}$$

$$\frac{\partial^2 Jns2\left(z_t\right)}{\partial z_t^2} = \left(\eta_{s,n}\right)^2 \frac{\partial^2 \pi\left[q_n^s\left(z_t\right)\right]}{\partial^2 q_n^s} - \left(\eta_{nc}\right)^2 \frac{\partial^2 \pi\left[q_{nc}\left(z_t\right)\right]}{\partial^2 q_{nc}\left(z_t\right)} \tag{83}$$

$$+\delta\rho^2 \frac{1}{n}\frac{\partial^2 V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial^2 z_{t+1}}$$

$$+\delta\frac{1}{n}n\eta_{s,n}\frac{\partial^2 V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial^2 z_{t+1}}\frac{dz_{t+1}}{dq_n^s}\frac{dq_n^s\left(z_t\right)}{dz_t} \tag{84}$$

$$+\delta\frac{1}{n}(N-n)\eta_{f,n}\frac{\partial^2 V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial^2 z_{t+1}}\frac{dz_{t+1}}{dq_n^f}\frac{dq_n^f\left(z_t\right)}{dz_t} \tag{85}$$

$$-\delta\rho^2\frac{1}{n}\frac{\partial^2 V_n^s\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]}{\partial^2 z_{t+1}} \tag{86}$$

$$-\delta\frac{1}{n}N\eta_{nc}\frac{\partial^2 V_n^s\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]}{\partial^2 z_{t+1}}\frac{dz_{t+1}}{dq_{nc}}\frac{dq_{nc}\left(z_t\right)}{dz_t} \tag{87}$$

$$\frac{\partial^2 Jns2\left(z_t\right)}{\partial z_t^2} = -\left(\eta_{s,n}\right)^2\alpha + \left(\eta_{nc}\right)^2\alpha$$

$$+\delta\rho^2\frac{1}{n}A_n^s$$

$$+\delta\frac{1}{n}n^2\left(\eta_{s,n}\right)^2 A_n^s + \delta\frac{1}{n}(N-n)^2\left(\eta_{f,n}\right)^2 A_n^s$$

$$-\delta\rho^2\frac{1}{n}A_n^s - \delta\frac{1}{n}N^2\left(\eta_{nc}\right)^2 A_n^s$$

$$\frac{\partial^2 ns2J\left(z_t\right)}{\partial z_t^2} = \alpha\left[\left(\eta_{nc}\right)^2 - \left(\eta_{s,n}\right)^2\right]$$

$$+A_n^s\left[\delta n\left(\eta_{s,n}\right)^2 + \delta\frac{1}{n}(N-n)^2\left(\eta_{f,n}^2\right)^2\right.$$

$$\left.-\delta\frac{1}{n}N^2\left(\eta_{nc}\right)^2\right]$$

$$\frac{\partial^2 Jns2\left(z_t\right)}{\partial z_t^2} = \alpha\left[\left(\eta_{nc}\right)^2 - \left(\eta_{s,n}\right)^2\right]$$

$$+A_n^s\delta\frac{1}{n}\left[n^2\left(\eta_{s,n}\right)^2 + (N-n)^2\left(\eta_{f,n}\right)^2\right.$$

$$\left.-N^2\left(\eta_{nc}\right)^2\right]$$

OK UP TO HERE

CHeck this -could perhaps get rid off the first two lines using the envelope theorem. Could there be some argument for why $A_n^s >$ or $< A_n^f$?

$$\frac{\partial \Phi\left(z_t\right)}{\partial z_t} = \left\{ \frac{\partial \pi\left[q_n^s\left(z_t\right)\right]}{\partial q_n^s} + \delta n \frac{\partial V_n^s\left[\rho z_t + n q_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial z_{t+1}} \right\} \frac{dq_n^s\left(z_t\right)}{dz_t} \tag{91}$$

$$-\left\{ \frac{\partial \pi\left[q_{n-1}^f\left(z_t\right)\right]}{\partial q_{n-1}^f\left(z_t\right)} + \delta \frac{\partial V_n^s\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]}{\partial z_{t+1}} \right\} \frac{dq_{n-1}^f\left(z_t\right)}{dz_t}$$

$$\tag{92}$$

$$+\delta\rho\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + n q_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial z_{t+1}} \tag{93}$$

$$-\delta(n-1)\frac{\partial V_n^s\left[\rho z_t + n q_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial z_{t+1}}\frac{dq_n^s\left(z_t\right)}{dz_t}$$

$$+\delta\frac{N-n}{n}\frac{\partial V_n^s\left[\rho z_t + n q_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial z_{t+1}}\frac{dq_n^f\left(z_t\right)}{dz_t}$$

$$-\delta\rho\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]}{\partial z_{t+1}}$$

$$-\delta\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]}{\partial z_{t+1}}\frac{dz_{t+1}}{dq_{n-1}^s}\frac{dq_{n-1}^s\left(z_t\right)}{dz_t}$$

$$-\delta\frac{(N-2n+1)}{n}\frac{\partial V_n^s\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]}{\partial z_{t+1}}\frac{dq_{n-1}^f\left(z_t\right)}{dz_t} \tag{94}$$

$$\tag{95}$$

$$\frac{\partial \Phi\left(z_t\right)}{\partial z_t} = \frac{\partial \pi\left[q_n^s\left(z_t\right)\right]}{\partial q_n^s}\frac{dq_n^s\left(z_t\right)}{dz_t} - \frac{\partial \pi\left[q_{n-1}^f\left(z_t\right)\right]}{\partial q_{n-1}^f\left(z_t\right)}\frac{dq_{n-1}^f\left(z_t\right)}{dz_t} \tag{96}$$

$$+\delta\rho\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + n q_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial z_{t+1}}$$

$$+\delta\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + n q_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial z_{t+1}}\frac{dz_{t+1}}{dq_n^s}\frac{dq_n^s\left(z_t\right)}{dz_t}$$

A signatory's payoff from deviating with "infinite punishment" is

$$J_{IP}\left(z_t\right) = \pi\left[q_n^s\left(z_t\right)\right] - \pi\left[q_{n-1}^f\left(z_t\right)\right]$$

$$+\delta\left[\frac{1}{n}V_n^s\left(z_t\right) - V_{n-1}^f\left(z_{t+1,n-1}\right)\right]$$

34

$$= \pi \left[q_n^s \left(z_t \right) \right] - \pi \left[q_{n-1}^f \left(z_t \right) \right]$$

$$+ \delta \frac{1}{n} V_n^s \left[\rho z_t + n q_n^s \left(z_t \right) + (N - n) q_n^f \left(z_t \right) \right]$$

$$- \delta V_n^f \left[\rho z_t + (n-1) q_{n-1}^s \left(z_t \right) + (N - n + 1) q_{n-1}^f \left(z_t \right) \right]$$

Differentiating with respect to z_t yieldsη

$$\frac{\partial J_{IP} \left(z_t \right)}{\partial z_t} = \frac{\partial \pi \left[q_n^s \left(z_t \right) \right]}{\partial q_n^s} \frac{d q_n^s \left(z_t \right)}{d z_t} - \frac{\partial \pi \left[q_{n-1}^f \left(z_t \right) \right]}{\partial q_{n-1}^f \left(z_t \right)} \frac{d q_{n-1}^f \left(z_t \right)}{d z_t}$$

$$+ \delta \rho \frac{1}{n} \frac{\partial V_n^s \left[\rho z_t + n q_n^s \left(z_t \right) + (N - n) q_n^f \left(z_t \right) \right]}{\partial z_{t+1}}$$

$$+ \delta \frac{1}{n} \frac{\partial V_n^s \left[\rho z_t + n q_n^s \left(z_t \right) + (N - n) q_n^f \left(z_t \right) \right]}{\partial z_{t+1}} \frac{d z_{t+1}}{d q_n^s} \frac{d q_n^s \left(z_t \right)}{d z_t}$$

$$+ \delta \frac{1}{n} \frac{\partial V_n^s \left[\rho z_t + n q_n^s \left(z_t \right) + (N - n) q_n^f \left(z_t \right) \right]}{\partial z_{t+1}} \frac{d z_{t+1}}{d q_n^f} \frac{d q_n^f \left(z_t \right)}{d z_t}$$

$$- \delta \rho \frac{\partial V_{n-1}^f \left[\rho z_t + (n-1) q_{n-1}^s \left(z_t \right) + (N - n + 1) q_{n-1}^f \left(z_t \right) \right]}{\partial z_{t+1}}$$

$$- \delta \frac{\partial V_{n-1}^f \left[\rho z_t + (n-1) q_{n-1}^s \left(z_t \right) + (N - n + 1) q_{n-1}^f \left(z_t \right) \right]}{\partial z_{t+1}} \frac{d z_{t+1}}{d q_{n-1}^s} \frac{d q_{n-1}^s \left(z_t \right)}{d z_t}$$

$$- \delta \frac{\partial V_{n-1}^f \left[\rho z_t + (n-1) q_{n-1}^s \left(z_t \right) + (N - n + 1) q_{n-1}^f \left(z_t \right) \right]}{\partial z_{t+1}} \frac{d z_{t+1}}{d q_{n-1}^f} \frac{d q_{n-1}^f \left(z_t \right)}{d z_t}$$

$$\frac{\partial J_{IP} \left(z_t \right)}{\partial z_t} = \frac{\partial \pi \left[q_n^s \left(z_t \right) \right]}{\partial q_n^s} \eta_{s,n} - \frac{\partial \pi \left[q_{n-1}^f \left(z_t \right) \right]}{\partial q_{n-1}^f \left(z_t \right)} \eta_{f,n-1} \tag{97}$$

$$+ \delta \rho \frac{1}{n} \frac{\partial V_n^s \left[\rho z_t + n q_n^s \left(z_t \right) + (N - n) q_n^f \left(z_t \right) \right]}{\partial z_{t+1}}$$

$$+ \delta \frac{1}{n} n \frac{\partial V_n^s \left[\rho z_t + n q_n^s \left(z_t \right) + (N - n) q_n^f \left(z_t \right) \right]}{\partial z_{t+1}} \eta_{s,n}$$

$$+ \delta \frac{1}{n} (N - n) \frac{\partial V_n^s \left[\rho z_t + n q_n^s \left(z_t \right) + (N - n) q_n^f \left(z_t \right) \right]}{\partial z_{t+1}} \eta_{f,n}$$

$$- \delta \rho \frac{\partial V_{n-1}^f \left[\rho z_t + (n-1) q_{n-1}^s \left(z_t \right) + (N - n + 1) q_{n-1}^f \left(z_t \right) \right]}{\partial z_{t+1}}$$

$$- \delta (n-1) \frac{\partial V_{n-1}^f \left[\rho z_t + (n-1) q_{n-1}^s \left(z_t \right) + (N - n + 1) q_{n-1}^f \left(z_t \right) \right]}{\partial z_{t+1}} \eta_{s,n-1}$$

$$- \delta (N - n + 1) \frac{\partial V_{n-1}^f \left[\rho z_t + (n-1) q_{n-1}^s \left(z_t \right) + (N - n + 1) q_{n-1}^f \left(z_t \right) \right]}{\partial z_{t+1}} \eta_{f,n-1}$$

$$\frac{\partial^2 J_{IP}(z_t)}{\partial z_t^2} = (\eta_{s,n})^2 \frac{\partial^2 \pi [q_n^s(z_t)]}{\partial^2 q_n^s} - (\eta_{f,n-1})^2 \frac{\partial \pi^2 \left[q_{n-1}^f(z_t)\right]}{\partial^2 q_{n-1}^f(z_t)} \tag{98}$$

$$+\delta \rho^2 \frac{1}{n} \frac{\partial^2 V_n^s \left[\rho z_t + n q_n^s(z_t) + (N-n) q_n^f(z_t)\right]}{\partial^2 z_{t+1}}$$

$$+\delta \frac{1}{n} n \eta_{s,n} \frac{\partial^2 V_n^s \left[\rho z_t + n q_n^s(z_t) + (N-n) q_n^f(z_t)\right]}{\partial^2 z_{t+1}} \frac{dz_{t+1}}{dq_n^s} \frac{dq_n^s(z_t)}{dz_t} \tag{99}$$

$$+\delta \frac{1}{n} (N-n) \eta_{f,n} \frac{\partial^2 V_n^s \left[\rho z_t + n q_n^s(z_t) + (N-n) q_n^f(z_t)\right]}{\partial^2 z_{t+1}} \frac{dz_{t+1}}{dq_n^f} \frac{dq_n^f(z_t)}{dz_t} \tag{100}$$

$$-\delta \rho^2 \frac{\partial^2 V_{n-1}^f \left[\rho z_t + (n-1) q_{n-1}^s(z_t) + (N-n+1) q_{n-1}^f(z_t)\right]}{\partial^2 z_{t+1}} \tag{101}$$

$$-\delta(n-1)\eta_{s,n-1} \frac{\partial^2 V_{n-1}^f \left[\rho z_t + (n-1) q_{n-1}^s(z_t) + (N-n+1) q_{n-1}^f(z_t)\right]}{\partial^2 z_{t+1}} \frac{dz_{t+1}}{dq_{n-1}^s} \frac{dq_{n-1}^s(z_t)}{dz_t}$$
$$\tag{102}$$

$$-\delta(N-n+1)\eta_{f,n-1} \frac{\partial V_{n-1}^f \left[\rho z_t + (n-1) q_{n-1}^s(z_t) + (N-n+1) q_{n-1}^f(z_t)\right]}{\partial z_{t+1}} \frac{dz_{t+1}}{dq_{n-1}^f} \frac{dq_{n-1}^f(z_t)}{dz_t}$$
$$\tag{103}$$

$$\frac{\partial^2 J_{IP}(z_t)}{\partial z_t^2} = -(\eta_{s,n})^2 \alpha + (\eta_{f,n-1})^2 \alpha$$

$$+\delta \frac{1}{n} A_n^s \rho^2$$

$$+\delta \frac{1}{n} A_n^s n^2 (\eta_{s,n})^2 + \delta \frac{1}{n} A_n^s (N-n)^2 \left(\eta_{f,n}^2\right)^2$$

$$-\delta \rho^2 A_{n-1}^f$$

$$-\delta(n-1)^2 (\eta_{s,n-1})^2 A_{n-1}^f - \delta(N-n+1)^2 (\eta_{f,n-1})^2 A_{n-1}^f$$

$$\frac{\partial^2 J_{IP}(z_t)}{\partial z_t^2} = -(\eta_{s,n})^2 \alpha + (\eta_{f,n-1})^2 \alpha$$

$$+\delta \frac{1}{n} A_n^s \left[\rho^2 + n^2 (\eta_{s,n})^2 + (N-n)^2 \left(\eta_{f,n}^2\right)^2\right] \tag{105a}$$

$$-\delta A_{n-1}^f \left[\rho^2 - (n-1)^2 (\eta_{s,n-1})^2 - (N-n+1)^2 (\eta_{f,n-1})^2\right]$$

OK UP TO HERE

CHeck this -could perhaps get rid off the first two lines using the envelope theorem.
Could there be some argument for why $A_n^s >$ or $< A_n^f$?

$$\frac{\partial \Phi\left(z_{t}\right)}{\partial z_{t}} = \left\{ \frac{\partial \pi\left[q_{n}^{s}\left(z_{t}\right)\right]}{\partial q_{n}^{s}} + \delta n \frac{\partial V_{n}^{s}\left[\rho z_{t} + nq_{n}^{s}\left(z_{t}\right) + (N-n)q_{n}^{f}\left(z_{t}\right)\right]}{\partial z_{t+1}} \right\} \frac{dq_{n}^{s}\left(z_{t}\right)}{dz_{t}} \qquad (106)$$

$$-\left\{ \frac{\partial \pi\left[q_{n-1}^{f}\left(z_{t}\right)\right]}{\partial q_{n-1}^{f}\left(z_{t}\right)} + \delta \frac{\partial V_{n}^{s}\left[\rho z_{t} + (n-1)q_{n-1}^{s}\left(z_{t}\right) + (N-n+1)q_{n-1}^{f}\left(z_{t}\right)\right]}{\partial z_{t+1}} \right\} \frac{dq_{n-1}^{f}\left(z_{t}\right)}{dz_{t}}$$

$$(107)$$

$$+\delta\rho\frac{1}{n}\frac{\partial V_{n}^{s}\left[\rho z_{t} + nq_{n}^{s}\left(z_{t}\right) + (N-n)q_{n}^{f}\left(z_{t}\right)\right]}{\partial z_{t+1}} \qquad (108)$$

$$-\delta(n-1)\frac{\partial V_{n}^{s}\left[\rho z_{t} + nq_{n}^{s}\left(z_{t}\right) + (N-n)q_{n}^{f}\left(z_{t}\right)\right]}{\partial z_{t+1}}\frac{dq_{n}^{s}\left(z_{t}\right)}{dz_{t}}$$

$$+\delta\frac{N-n}{n}\frac{\partial V_{n}^{s}\left[\rho z_{t} + nq_{n}^{s}\left(z_{t}\right) + (N-n)q_{n}^{f}\left(z_{t}\right)\right]}{\partial z_{t+1}}\frac{dq_{n}^{f}\left(z_{t}\right)}{dz_{t}}$$

$$-\delta\rho\frac{1}{n}\frac{\partial V_{n}^{s}\left[\rho z_{t} + (n-1)q_{n-1}^{s}\left(z_{t}\right) + (N-n+1)q_{n-1}^{f}\left(z_{t}\right)\right]}{\partial z_{t+1}}$$

$$-\delta\frac{1}{n}\frac{\partial V_{n}^{s}\left[\rho z_{t} + (n-1)q_{n-1}^{s}\left(z_{t}\right) + (N-n+1)q_{n-1}^{f}\left(z_{t}\right)\right]}{\partial z_{t+1}}\frac{dz_{t+1}}{dq_{n-1}^{s}}\frac{dq_{n-1}^{s}\left(z_{t}\right)}{dz_{t}}$$

$$-\delta\frac{(N-2n+1)}{n}\frac{\partial V_{n}^{s}\left[\rho z_{t} + (n-1)q_{n-1}^{s}\left(z_{t}\right) + (N-n+1)q_{n-1}^{f}\left(z_{t}\right)\right]}{\partial z_{t+1}}\frac{dq_{n-1}^{f}\left(z_{t}\right)}{dz_{t}} \qquad (109)$$

$$(110)$$

$$\frac{\partial \Phi\left(z_{t}\right)}{\partial z_{t}} = \frac{\partial \pi\left[q_{n}^{s}\left(z_{t}\right)\right]}{\partial q_{n}^{s}}\frac{dq_{n}^{s}\left(z_{t}\right)}{dz_{t}} - \frac{\partial \pi\left[q_{n-1}^{f}\left(z_{t}\right)\right]}{\partial q_{n-1}^{f}\left(z_{t}\right)}\frac{dq_{n-1}^{f}\left(z_{t}\right)}{dz_{t}} \qquad (111)$$

$$+\delta\rho\frac{1}{n}\frac{\partial V_{n}^{s}\left[\rho z_{t} + nq_{n}^{s}\left(z_{t}\right) + (N-n)q_{n}^{f}\left(z_{t}\right)\right]}{\partial z_{t+1}}$$

$$+\delta\frac{1}{n}\frac{\partial V_{n}^{s}\left[\rho z_{t} + nq_{n}^{s}\left(z_{t}\right) + (N-n)q_{n}^{f}\left(z_{t}\right)\right]}{\partial z_{t+1}}\frac{dz_{t+1}}{dq_{n}^{s}}\frac{dq_{n}^{s}\left(z_{t}\right)}{dz_{t}}$$

A signatory's payoff from deviating at ns=2 (with return to cooperation in period $t+1$) is

$$Jns2\left(z_{t}\right) = \pi\left[q_{n}^{s}\left(z_{t}\right)\right] - \pi\left[q_{nc}\left(z_{t}\right)\right]$$

$$+\delta\frac{1}{n}\left[V_{n}^{s}\left(z_{t}\right) - V_{n}^{s}\left(z_{t+1,n-1}\right)\right]$$

with

$$z_{t+1,n-1} = \rho z_t + N q_{nc}(z_t)$$

$$Jns2(z_t) = \pi[q_n^s(z_t)] - \pi[q_{nc}(z_t)]$$
$$+\delta \frac{1}{n} V_n^s [\rho z_t + n q_n^s(z_t) + (N-n)q_n^f(z_t)]$$
$$-\delta \frac{1}{n} V_n^s [\rho z_t + N q_{nc}(z_t)]$$

Differentiating with respect to z_t yieldsη

$$\frac{\partial ns2J(z_t)}{\partial z_t} = \frac{\partial \pi[q_n^s(z_t)]}{\partial q_n^s} \frac{dq_n^s(z_t)}{dz_t} - \frac{\partial \pi[q_{nc}(z_t)]}{\partial q_{nc}(z_t)} \frac{dq_{nc}(z_t)}{dz_t} \tag{112}$$
$$+\delta \rho \frac{1}{n} \frac{\partial V_n^s [\rho z_t + n q_n^s(z_t) + (N-n)q_n^f(z_t)]}{\partial z_{t+1}}$$
$$+\delta \frac{1}{n} \frac{\partial V_n^s [\rho z_t + n q_n^s(z_t) + (N-n)q_n^f(z_t)]}{\partial z_{t+1}} \frac{dz_{t+1}}{dq_n^s} \frac{dq_n^s(z_t)}{dz_t}$$
$$+\delta \frac{1}{n} \frac{\partial V_n^s [\rho z_t + n q_n^s(z_t) + (N-n)q_n^f(z_t)]}{\partial z_{t+1}} \frac{dz_{t+1}}{dq_n^f} \frac{dq_n^f(z_t)}{dz_t}$$
$$-\delta \rho \frac{1}{n} \frac{\partial V_n^s [\rho z_t + (n-1)q_{n-1}^s(z_t) + (N-n+1)q_{n-1}^f(z_t)]}{\partial z_{t+1}}$$
$$-\delta \frac{1}{n} \frac{\partial V_n^s [\rho z_t + (n-1)q_{n-1}^s(z_t) + (N-n+1)q_{n-1}^f(z_t)]}{\partial z_{t+1}} \frac{dz_{t+1}}{dq_{nc}} \frac{dq_{nc}(z_t)}{dz_t}$$

$$\frac{\partial Jns2(z_t)}{\partial z_t} = \frac{\partial \pi[q_n^s(z_t)]}{\partial q_n^s} \eta_{s,n} - \frac{\partial \pi[q_{nc}(z_t)]}{\partial q_{nc}(z_t)} \eta_{nc} \tag{113}$$
$$+\delta \rho \frac{1}{n} \frac{\partial V_n^s [\rho z_t + n q_n^s(z_t) + (N-n)q_n^f(z_t)]}{\partial z_{t+1}}$$
$$+\delta \frac{1}{n} n \frac{\partial V_n^s [\rho z_t + n q_n^s(z_t) + (N-n)q_n^f(z_t)]}{\partial z_{t+1}} \eta_{s,n}$$
$$+\delta \frac{1}{n} (N-n) \frac{\partial V_n^s [\rho z_t + n q_n^s(z_t) + (N-n)q_n^f(z_t)]}{\partial z_{t+1}} \eta_{f,n}$$
$$-\delta \rho \frac{1}{n} \frac{\partial V_n^s [\rho z_t + N q_{nc}(z_t)]}{\partial z_{t+1}}$$
$$-\delta \frac{1}{n} N \frac{\partial V_n^s [\rho z_t + N q_{nc}(z_t)]}{\partial z_{t+1}} \eta_{nc}$$

$$\frac{\partial^2 Jns2\,(z_t)}{\partial z_t^2} = (\eta_{s,n})^2 \frac{\partial^2 \pi \left[q_n^s\,(z_t)\right]}{\partial^2 q_n^s} - (\eta_{nc})^2 \frac{\partial^2 \pi \left[q_{nc}\,(z_t)\right]}{\partial^2 q_{nc}\,(z_t)} \qquad (114)$$

$$+ \delta \rho^2 \frac{1}{n} \frac{\partial^2 V_n^s \left[\rho z_t + n q_n^s\,(z_t) + (N-n) q_n^f\,(z_t)\right]}{\partial^2 z_{t+1}}$$

$$+ \delta \frac{1}{n} n \eta_{s,n} \frac{\partial^2 V_n^s \left[\rho z_t + n q_n^s\,(z_t) + (N-n) q_n^f\,(z_t)\right]}{\partial^2 z_{t+1}} \frac{dz_{t+1}}{dq_n^s} \frac{dq_n^s\,(z_t)}{dz_t} \qquad (115)$$

$$+ \delta \frac{1}{n} (N-n) \eta_{f,n} \frac{\partial^2 V_n^s \left[\rho z_t + n q_n^s\,(z_t) + (N-n) q_n^f\,(z_t)\right]}{\partial^2 z_{t+1}} \frac{dz_{t+1}}{dq_n^f} \frac{dq_n^f\,(z_t)}{dz_t} \qquad (116)$$

$$- \delta \rho^2 \frac{1}{n} \frac{\partial^2 V_n^s \left[\rho z_t + (n-1) q_{n-1}^s\,(z_t) + (N-n+1) q_{n-1}^f\,(z_t)\right]}{\partial^2 z_{t+1}} \qquad (117)$$

$$- \delta \frac{1}{n} N \eta_{nc} \frac{\partial^2 V_n^s \left[\rho z_t + (n-1) q_{n-1}^s\,(z_t) + (N-n+1) q_{n-1}^f\,(z_t)\right]}{\partial^2 z_{t+1}} \frac{dz_{t+1}}{dq_{nc}} \frac{dq_{nc}\,(z_t)}{dz_t} \qquad (118)$$

$$\frac{\partial^2 Jns2\,(z_t)}{\partial z_t^2} = - (\eta_{s,n})^2 \alpha + (\eta_{nc})^2 \alpha$$

$$+ \delta \rho^2 \frac{1}{n} A_n^s$$

$$+ \delta \frac{1}{n} n^2 (\eta_{s,n})^2 A_n^s + \delta \frac{1}{n} (N-n)^2 (\eta_{f,n})^2 A_n^s$$

$$- \delta \rho^2 \frac{1}{n} A_n^s - \delta \frac{1}{n} N^2 (\eta_{nc})^2 A_n^s$$

$$\frac{\partial^2 ns2J\,(z_t)}{\partial z_t^2} = \alpha \left[(\eta_{nc})^2 - (\eta_{s,n})^2\right]$$

$$+ A_n^s \left[\delta n (\eta_{s,n})^2 + \delta \frac{1}{n} (N-n)^2 \left(\eta_{f,n}^2\right)^2\right.$$

$$\left. - \delta \frac{1}{n} N^2 (\eta_{nc})^2\right]$$

$$\frac{\partial^2 Jns2\,(z_t)}{\partial z_t^2} = \alpha \left[(\eta_{nc})^2 - (\eta_{s,n})^2\right]$$

$$+ A_n^s \delta \frac{1}{n} \left[n^2 (\eta_{s,n})^2 + (N-n)^2 (\eta_{f,n})^2\right.$$

$$\left. - N^2 (\eta_{nc})^2\right]$$

A signatory's payoff from deviating with "infinite punishment" and $n = 2$ is

$$J_{IPn2}(z_t) = \pi\left[q_n^s(z_t)\right] - \pi\left[q^{nc}(z_t)\right]$$

$$+\delta\left[\frac{1}{n}V_n^s(z_t) - V^{nc}(z_{t+1})\right]$$

$$= \pi\left[q_n^s(z_t)\right] - \pi\left[q^{nc}(z_t)\right]$$

$$+\delta\frac{1}{n}V_n^s\left[\rho z_t + nq_n^s(z_t) + (N-n)q_n^f(z_t)\right]$$

$$-\delta V^{nc}\left[\rho z_t + Nq^{nc}(z_t)\right]$$

Differentiating with respect to z_t yieldsη

$$\frac{\partial J_{IPn2}(z_t)}{\partial z_t} = \frac{\partial\pi\left[q_n^s(z_t)\right]}{\partial q_n^s}\frac{dq_n^s(z_t)}{dz_t} - \frac{\partial\pi\left[q^{nc}(z_t)\right]}{\partial q^{nc}(z_t)}\frac{dq^{nc}(z_t)}{dz_t}$$

$$+\delta\rho\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + nq_n^s(z_t) + (N-n)q_n^f(z_t)\right]}{\partial z_{t+1}}$$

$$+\delta\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + nq_n^s(z_t) + (N-n)q_n^f(z_t)\right]}{\partial z_{t+1}}\frac{dz_{t+1}}{dq_n^s}\frac{dq_n^s(z_t)}{dz_t}$$

$$+\delta\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + nq_n^s(z_t) + (N-n)q_n^f(z_t)\right]}{\partial z_{t+1}}\frac{dz_{t+1}}{dq_n^f}\frac{dq_n^f(z_t)}{dz_t}$$

$$-\delta\rho\frac{\partial V^{nc}\left[\rho z_t + Nq^{nc}(z_t)\right]}{\partial z_{t+1}}$$

$$-\delta\frac{\partial V^{nc}\left[\rho z_t + Nq^{nc}(z_t)\right]}{\partial z_{t+1}}\frac{dz_{t+1}}{dq^{nc}}\frac{dq^{nc}(z_t)}{dz_t}$$

$$\frac{\partial J_{IPn2}(z_t)}{\partial z_t} = \frac{\partial\pi\left[q_n^s(z_t)\right]}{\partial q_n^s}\eta_{s,n} - \frac{\partial\pi\left[q^{nc}(z_t)\right]}{\partial q^{nc}(z_t)}\eta_{nc}$$

$$+\delta\rho\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + nq_n^s(z_t) + (N-n)q_n^f(z_t)\right]}{\partial z_{t+1}}$$

$$+\delta\frac{1}{n}n\frac{\partial V_n^s\left[\rho z_t + nq_n^s(z_t) + (N-n)q_n^f(z_t)\right]}{\partial z_{t+1}}\eta_{s,n}$$

$$+\delta\frac{1}{n}(N-n)\frac{\partial V_n^s\left[\rho z_t + nq_n^s(z_t) + (N-n)q_n^f(z_t)\right]}{\partial z_{t+1}}\eta_{f,n}$$

$$-\delta\rho\frac{\partial V^{nc}\left[\rho z_t + Nq^{nc}(z_t)\right]}{\partial z_{t+1}}$$

$$-\delta N\frac{\partial V^{nc}\left[\rho z_t + Nq^{nc}(z_t)\right]}{\partial z_{t+1}}\eta_{nc}$$

$$\frac{\partial^2 J_{IPn2}(z_t)}{\partial z_t^2} = (\eta_{s,n})^2 \frac{\partial^2 \pi \left[q_n^s(z_t) \right]}{\partial^2 q_n^s} - (\eta_{nc})^2 \frac{\partial \pi^2 \left[q^{nc}(z_t) \right]}{\partial^2 q^{nc}(z_t)}$$

$$+ \delta \rho^2 \frac{1}{n} \frac{\partial^2 V_n^s \left[\rho z_t + n q_n^s(z_t) + (N-n) q_n^f(z_t) \right]}{\partial^2 z_{t+1}}$$

$$+ \delta \frac{1}{n} n \eta_{s,n} \frac{\partial^2 V_n^s \left[\rho z_t + n q_n^s(z_t) + (N-n) q_n^f(z_t) \right]}{\partial^2 z_{t+1}} \frac{dz_{t+1}}{dq_n^s} \frac{dq_n^s(z_t)}{dz_t} \quad (122)$$

$$+ \delta \frac{1}{n}(N-n)\eta_{f,n} \frac{\partial^2 V_n^s \left[\rho z_t + n q_n^s(z_t) + (N-n) q_n^f(z_t) \right]}{\partial^2 z_{t+1}} \frac{dz_{t+1}}{dq_n^f} \frac{dq_n^f(z_t)}{dz_t}$$

$$- \delta \rho^2 \frac{\partial^2 V^{nc} \left[\rho z_t + N q^{nc}(z_t) \right]}{\partial^2 z_{t+1}}$$

$$- \delta N \eta_{nc} \frac{\partial^2 V^{nc} \left[\rho z_t + N q^{nc}(z_t) \right]}{\partial^2 z_{t+1}} \frac{dz_{t+1}}{dq^{nc}} \frac{dq^{nc}(z_t)}{dz_t}$$

$$\frac{\partial^2 J_{IP}(z_t)}{\partial z_t^2} = -(\eta_{s,n})^2 \alpha + (\eta_{f,n-1})^2 \alpha$$

$$+ \delta \frac{1}{n} A_n^s \rho^2$$

$$+ \delta \frac{1}{n} A_n^s n^2 (\eta_{s,n})^2 + \delta \frac{1}{n} A_n^s (N-n)^2 \left(\eta_{f,n}^2 \right)^2$$

$$- \delta \rho^2 A_{n-1}^f$$

$$- \delta (n-1)^2 (\eta_{s,n-1})^2 A_{n-1}^f - \delta (N-n+1)^2 (\eta_{f,n-1})^2 A_{n-1}^f$$

$$\frac{\partial^2 J_{IP}(z_t)}{\partial z_t^2} = -(\eta_{s,n})^2 \alpha + (\eta_{f,n-1})^2 \alpha$$

$$+ \delta \frac{1}{n} A_n^s \left[\rho^2 + n^2 (\eta_{s,n})^2 + (N-n)^2 \left(\eta_{f,n}^2 \right)^2 \right] \quad (124a)$$

$$- \delta A_{n-1}^f \left[\rho^2 - (n-1)^2 (\eta_{s,n-1})^2 - (N-n+1)^2 (\eta_{f,n-1})^2 \right]$$

OK UP TO HERE

CHeck this -could perhaps get rid off the first two lines using the envelope theorem. Could there be some argument for why $A_n^s >$ or $< A_n^f$?

$$\frac{\partial \Phi\left(z_{t}\right)}{\partial z_{t}} = \left\{\frac{\partial \pi\left[q_{n}^{s}\left(z_{t}\right)\right]}{\partial q_{n}^{s}} + \delta n \frac{\partial V_{n}^{s}\left[\rho z_{t} + n q_{n}^{s}\left(z_{t}\right) + (N-n) q_{n}^{f}\left(z_{t}\right)\right]}{\partial z_{t+1}}\right\} \frac{d q_{n}^{s}\left(z_{t}\right)}{d z_{t}} \tag{125}$$

$$-\left\{\frac{\partial \pi\left[q_{n-1}^{f}\left(z_{t}\right)\right]}{\partial q_{n-1}^{f}\left(z_{t}\right)} + \delta \frac{\partial V_{n}^{s}\left[\rho z_{t} + (n-1) q_{n-1}^{s}\left(z_{t}\right) + (N-n+1) q_{n-1}^{f}\left(z_{t}\right)\right]}{\partial z_{t+1}}\right\} \frac{d q_{n-1}^{f}\left(z_{t}\right)}{d z_{t}}$$

$$\tag{126}$$

$$+\delta \rho \frac{1}{n} \frac{\partial V_{n}^{s}\left[\rho z_{t} + n q_{n}^{s}\left(z_{t}\right) + (N-n) q_{n}^{f}\left(z_{t}\right)\right]}{\partial z_{t+1}} \tag{127}$$

$$-\delta(n-1) \frac{\partial V_{n}^{s}\left[\rho z_{t} + n q_{n}^{s}\left(z_{t}\right) + (N-n) q_{n}^{f}\left(z_{t}\right)\right]}{\partial z_{t+1}} \frac{d q_{n}^{s}\left(z_{t}\right)}{d z_{t}}$$

$$+\delta \frac{N-n}{n} \frac{\partial V_{n}^{s}\left[\rho z_{t} + n q_{n}^{s}\left(z_{t}\right) + (N-n) q_{n}^{f}\left(z_{t}\right)\right]}{\partial z_{t+1}} \frac{d q_{n}^{f}\left(z_{t}\right)}{d z_{t}}$$

$$-\delta \rho \frac{1}{n} \frac{\partial V_{n}^{s}\left[\rho z_{t} + (n-1) q_{n-1}^{s}\left(z_{t}\right) + (N-n+1) q_{n-1}^{f}\left(z_{t}\right)\right]}{\partial z_{t+1}}$$

$$-\delta \frac{1}{n} \frac{\partial V_{n}^{s}\left[\rho z_{t} + (n-1) q_{n-1}^{s}\left(z_{t}\right) + (N-n+1) q_{n-1}^{f}\left(z_{t}\right)\right]}{\partial z_{t+1}} \frac{d z_{t+1}}{d q_{n-1}^{s}} \frac{d q_{n-1}^{s}\left(z_{t}\right)}{d z_{t}}$$

$$-\delta \frac{(N-2n+1)}{n} \frac{\partial V_{n}^{s}\left[\rho z_{t} + (n-1) q_{n-1}^{s}\left(z_{t}\right) + (N-n+1) q_{n-1}^{f}\left(z_{t}\right)\right]}{\partial z_{t+1}} \frac{d q_{n-1}^{f}\left(z_{t}\right)}{d z_{t}} \tag{128}$$

$$\tag{129}$$

$$\frac{\partial \Phi\left(z_{t}\right)}{\partial z_{t}} = \frac{\partial \pi\left[q_{n}^{s}\left(z_{t}\right)\right]}{\partial q_{n}^{s}} \frac{d q_{n}^{s}\left(z_{t}\right)}{d z_{t}} - \frac{\partial \pi\left[q_{n-1}^{f}\left(z_{t}\right)\right]}{\partial q_{n-1}^{f}\left(z_{t}\right)} \frac{d q_{n-1}^{f}\left(z_{t}\right)}{d z_{t}} \tag{130}$$

$$+\delta \rho \frac{1}{n} \frac{\partial V_{n}^{s}\left[\rho z_{t} + n q_{n}^{s}\left(z_{t}\right) + (N-n) q_{n}^{f}\left(z_{t}\right)\right]}{\partial z_{t+1}}$$

$$+\delta \frac{1}{n} \frac{\partial V_{n}^{s}\left[\rho z_{t} + n q_{n}^{s}\left(z_{t}\right) + (N-n) q_{n}^{f}\left(z_{t}\right)\right]}{\partial z_{t+1}} \frac{d z_{t+1}}{d q_{n}^{s}} \frac{d q_{n}^{s}\left(z_{t}\right)}{d z_{t}}$$

A signatory's payoff from deviating at ns=2 (with return to cooperation in period $t+1$) is

$$Jns2\left(z_{t}\right) = \pi\left[q_{n}^{s}\left(z_{t}\right)\right] - \pi\left[q_{nc}\left(z_{t}\right)\right]$$

$$+\delta \frac{1}{n}\left[V_{n}^{s}\left(z_{t}\right) - V_{n}^{s}\left(z_{t+1,n-1}\right)\right]$$

with

$$z_{t+1,n-1} = \rho z_t + N q_{nc}(z_t)$$

$$Jns2(z_t) = \pi\left[q_n^s(z_t)\right] - \pi\left[q_{nc}(z_t)\right]$$
$$+\delta\frac{1}{n}V_n^s\left[\rho z_t + n q_n^s(z_t) + (N-n)q_n^f(z_t)\right]$$
$$-\delta\frac{1}{n}V_n^s\left[\rho z_t + N q_{nc}(z_t)\right]$$

Differentiating with respect to z_t yieldsη

$$\frac{\partial ns2J(z_t)}{\partial z_t} = \frac{\partial\pi\left[q_n^s(z_t)\right]}{\partial q_n^s}\frac{dq_n^s(z_t)}{dz_t} - \frac{\partial\pi\left[q_{nc}(z_t)\right]}{\partial q_{nc}(z_t)}\frac{dq_{nc}(z_t)}{dz_t} \qquad (131)$$
$$+\delta\rho\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + n q_n^s(z_t) + (N-n)q_n^f(z_t)\right]}{\partial z_{t+1}}$$
$$+\delta\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + n q_n^s(z_t) + (N-n)q_n^f(z_t)\right]}{\partial z_{t+1}}\frac{dz_{t+1}}{dq_n^s}\frac{dq_n^s(z_t)}{dz_t}$$
$$+\delta\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + n q_n^s(z_t) + (N-n)q_n^f(z_t)\right]}{\partial z_{t+1}}\frac{dz_{t+1}}{dq_n^f}\frac{dq_n^f(z_t)}{dz_t}$$
$$-\delta\rho\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + (n-1)q_{n-1}^s(z_t) + (N-n+1)q_{n-1}^f(z_t)\right]}{\partial z_{t+1}}$$
$$-\delta\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + (n-1)q_{n-1}^s(z_t) + (N-n+1)q_{n-1}^f(z_t)\right]}{\partial z_{t+1}}\frac{dz_{t+1}}{dq_{nc}}\frac{dq_{nc}(z_t)}{dz_t}$$

$$\frac{\partial Jns2(z_t)}{\partial z_t} = \frac{\partial\pi\left[q_n^s(z_t)\right]}{\partial q_n^s}\eta_{s,n} - \frac{\partial\pi\left[q_{nc}(z_t)\right]}{\partial q_{nc}(z_t)}\eta_{nc} \qquad (132)$$
$$+\delta\rho\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + n q_n^s(z_t) + (N-n)q_n^f(z_t)\right]}{\partial z_{t+1}}$$
$$+\delta\frac{1}{n}n\frac{\partial V_n^s\left[\rho z_t + n q_n^s(z_t) + (N-n)q_n^f(z_t)\right]}{\partial z_{t+1}}\eta_{s,n}$$
$$+\delta\frac{1}{n}(N-n)\frac{\partial V_n^s\left[\rho z_t + n q_n^s(z_t) + (N-n)q_n^f(z_t)\right]}{\partial z_{t+1}}\eta_{f,n}$$
$$-\delta\rho\frac{1}{n}\frac{\partial V_n^s\left[\rho z_t + N q_{nc}(z_t)\right]}{\partial z_{t+1}}$$
$$-\delta\frac{1}{n}N\frac{\partial V_n^s\left[\rho z_t + N q_{nc}(z_t)\right]}{\partial z_{t+1}}\eta_{nc}$$

$$\frac{\partial^2 Jns2\left(z_t\right)}{\partial z_t^2} = \left(\eta_{s,n}\right)^2 \frac{\partial^2 \pi\left[q_n^s\left(z_t\right)\right]}{\partial^2 q_n^s} - \left(\eta_{nc}\right)^2 \frac{\partial^2 \pi\left[q_{nc}\left(z_t\right)\right]}{\partial^2 q_{nc}\left(z_t\right)} \tag{133}$$

$$+\delta\rho^2 \frac{1}{n} \frac{\partial^2 V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial^2 z_{t+1}}$$

$$+\delta\frac{1}{n}n\eta_{s,n} \frac{\partial^2 V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial^2 z_{t+1}} \frac{dz_{t+1}}{dq_n^s} \frac{dq_n^s\left(z_t\right)}{dz_t} \tag{134}$$

$$+\delta\frac{1}{n}(N-n)\eta_{f,n} \frac{\partial^2 V_n^s\left[\rho z_t + nq_n^s\left(z_t\right) + (N-n)q_n^f\left(z_t\right)\right]}{\partial^2 z_{t+1}} \frac{dz_{t+1}}{dq_n^f} \frac{dq_n^f\left(z_t\right)}{dz_t} \tag{135}$$

$$-\delta\rho^2 \frac{1}{n} \frac{\partial^2 V_n^s\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]}{\partial^2 z_{t+1}} \tag{136}$$

$$-\delta\frac{1}{n}N\eta_{nc} \frac{\partial^2 V_n^s\left[\rho z_t + (n-1)q_{n-1}^s\left(z_t\right) + (N-n+1)q_{n-1}^f\left(z_t\right)\right]}{\partial^2 z_{t+1}} \frac{dz_{t+1}}{dq_{nc}} \frac{dq_{nc}\left(z_t\right)}{dz_t} \tag{137}$$

$$\frac{\partial^2 Jns2\left(z_t\right)}{\partial z_t^2} = -\left(\eta_{s,n}\right)^2 \alpha + \left(\eta_{nc}\right)^2 \alpha$$

$$+\delta\rho^2 \frac{1}{n}A_n^s$$

$$+\delta\frac{1}{n}n^2\left(\eta_{s,n}\right)^2 A_n^s + \delta\frac{1}{n}(N-n)^2\left(\eta_{f,n}\right)^2 A_n^s$$

$$-\delta\rho^2 \frac{1}{n}A_n^s - \delta\frac{1}{n}N^2\left(\eta_{nc}\right)^2 A_n^s$$

$$\frac{\partial^2 ns2J\left(z_t\right)}{\partial z_t^2} = \alpha\left[\left(\eta_{nc}\right)^2 - \left(\eta_{s,n}\right)^2\right]$$

$$+A_n^s\left[\delta n\left(\eta_{s,n}\right)^2 + \delta\frac{1}{n}(N-n)^2\left(\eta_{f,n}^2\right)^2\right.$$

$$\left.-\delta\frac{1}{n}N^2\left(\eta_{nc}\right)^2\right]$$

$$\frac{\partial^2 Jns2\left(z_t\right)}{\partial z_t^2} = \alpha\left[\left(\eta_{nc}\right)^2 - \left(\eta_{s,n}\right)^2\right]$$

$$+A_n^s\delta\frac{1}{n}\left[n^2\left(\eta_{s,n}\right)^2 + (N-n)^2\left(\eta_{f,n}\right)^2\right.$$

$$\left.-N^2\left(\eta_{nc}\right)^2\right]$$

OK UP TO HERE

CHeck this -could perhaps get rid off the first two lines using the envelope theorem. Could there be some argument for why $A_n^s >$ or $< A_n^f$?

$$\frac{\partial \Phi(z_t)}{\partial z_t} = \left\{ \frac{\partial \pi \left[q_n^s(z_t) \right]}{\partial q_n^s} + \delta n \frac{\partial V_n^s \left[\rho z_t + n q_n^s(z_t) + (N-n) q_n^f(z_t) \right]}{\partial z_{t+1}} \right\} \frac{dq_n^s(z_t)}{dz_t} \quad (141)$$

$$- \left\{ \frac{\partial \pi \left[q_{n-1}^f(z_t) \right]}{\partial q_{n-1}^f(z_t)} + \delta \frac{\partial V_n^s \left[\rho z_t + (n-1) q_{n-1}^s(z_t) + (N-n+1) q_{n-1}^f(z_t) \right]}{\partial z_{t+1}} \right\} \frac{dq_{n-1}^f(z_t)}{dz_t}$$

$$(142)$$

$$+ \delta \rho \frac{1}{n} \frac{\partial V_n^s \left[\rho z_t + n q_n^s(z_t) + (N-n) q_n^f(z_t) \right]}{\partial z_{t+1}} \quad (143)$$

$$- \delta(n-1) \frac{\partial V_n^s \left[\rho z_t + n q_n^s(z_t) + (N-n) q_n^f(z_t) \right]}{\partial z_{t+1}} \frac{dq_n^s(z_t)}{dz_t}$$

$$+ \delta \frac{N-n}{n} \frac{\partial V_n^s \left[\rho z_t + n q_n^s(z_t) + (N-n) q_n^f(z_t) \right]}{\partial z_{t+1}} \frac{dq_n^f(z_t)}{dz_t}$$

$$- \delta \rho \frac{1}{n} \frac{\partial V_n^s \left[\rho z_t + (n-1) q_{n-1}^s(z_t) + (N-n+1) q_{n-1}^f(z_t) \right]}{\partial z_{t+1}}$$

$$- \delta \frac{1}{n} \frac{\partial V_n^s \left[\rho z_t + (n-1) q_{n-1}^s(z_t) + (N-n+1) q_{n-1}^f(z_t) \right]}{\partial z_{t+1}} \frac{dz_{t+1}}{dq_{n-1}^s} \frac{dq_{n-1}^s(z_t)}{dz_t}$$

$$- \delta \frac{(N-2n+1)}{n} \frac{\partial V_n^s \left[\rho z_t + (n-1) q_{n-1}^s(z_t) + (N-n+1) q_{n-1}^f(z_t) \right]}{\partial z_{t+1}} \frac{dq_{n-1}^f(z_t)}{dz_t} \quad (144)$$

$$(145)$$