April 2003

Tough Love: Optimal Enforcement of Output Quotas in the Presence of Cheating

Konstantinos Giannakas
Department of Agricultural Economics, University of Nebraska-Lincoln, Lincoln, Nebraska, USA, kgiannakas2@unl.edu

Murray Fulton
University of Saskatchewan, Saskatoon, SK, Canada

Follow this and additional works at: http://digitalcommons.unl.edu/ageconfacpub
Part of the Agricultural and Resource Economics Commons

http://digitalcommons.unl.edu/ageconfacpub/39

This Article is brought to you for free and open access by the Agricultural Economics Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications: Agricultural Economics by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Tough Love: Optimal Enforcement of Output Quotas in the Presence of Cheating

Konstantinos Giannakas* Murray Fulton†

*Department of Agricultural Economics, University of Nebraska-Lincoln, Lincoln, Nebraska, USA,
†Department of Agricultural Economics, University of Saskatchewan, Saskatoon, SK, Canada,

Copyright ©2003 by the authors. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher, bepress, which has been given certain exclusive rights by the author. Journal of Agricultural & Food Industrial Organization is produced by The Berkeley Electronic Press (bepress). http://www.bepress.com/jafio
Tough Love: Optimal Enforcement of Output Quotas in the Presence of Cheating

Konstantinos Giannakas and Murray Fulton

Abstract

This study builds on the literature on the economics of output quotas in the presence of cheating. We extend previous work by Giannakas and Fulton (2000a) by examining the decisions of the agency responsible for the enforcement of output quotas in a decentralized policy-making structure. Enforcement policy design is modeled as a sequential game between the agency that determines the enforcement of the quota program (designed by a regulator), and farmers who make production decisions. Analytical results show that the level of enforcement depends on the size of the enforcement costs and the political preferences of the enforcement agency - the greater is the weight placed by the enforcement agency on producer welfare, the greater is the level of program enforcement.
1. Introduction

Governments have used a variety of market intervention mechanisms to redistribute income in the economy. Output quotas have been widely used in a number of different areas including agriculture and fisheries. In agriculture, where quotas have been extensively analyzed, examples of supply restrictions include the tobacco and peanut programs in the United States (US), milk quotas in the European Union (EU), and supply management of milk, chicken and eggs in Canada.¹

Quotas redistribute income by restricting supply of the regulated commodity, thus raising its price and generating higher returns for the original quota holders. This increased price however, provides an incentive for producers to expand production beyond the restricted amount. And this supply expansion is what is generally observed. Again with reference to agriculture, violation of quota limits and excess production of regulated commodities are common in the EU Common Agricultural Policy (CAP) (Ockenden and Franklin, 1995; Gardner, 1996). This is especially true for the milk quotas in the Mediterranean regions. In 1997, the penalties on detected above-quota production of milk in Italy alone exceeded $480 million (Tagliabue, 1998). Alston (1986) reports that when hen quotas were used to control the egg market in the state of Victoria, Australia, the black market was estimated to account for 10-30% of all eggs. In contrast, over-quota chicken production in Canada is not rampant, although it is significant from time to time in various areas of the country.

Despite the incentive for and the incidence of cheating on output quotas, this issue has received limited attention in the economics literature. The traditional analysis of this policy takes place under the unrealistic assumption of perfect and costless enforcement. An exception is the paper by Giannakas and Fulton (2000a) that introduces enforcement costs and cheating into the economic analysis of the income transfer efficiency of quotas.²

In this paper, we extend the work of Giannakas and Fulton (2000a, GF hereafter) to include the optimal enforcement response to producer noncompliant behavior. In particular, unlike GF that examine the transfer efficiency of output quotas under different levels of exogenously determined policy enforcement, this paper explicitly considers the optimal degree of enforcement (i.e., the problem of the agency responsible for the enforcement of the quota program) when enforcement is costly and imperfect.

In addition to extending the GF study to the analysis of the optimal enforcement policy in the presence of enforcement costs and farmer noncompliant behavior, this paper incorporates a much richer (and more realistic) policy making structure. In particular, GF analyzes the transfer efficiency effects of quota violations in a centralized policy making structure where policy design and implementation are the responsibility of a single agency (the government). This concentration of policy functions is rarely the case, however.

The agricultural policy making structure in most developed countries, and certainly the US, the EU and Canada, involves a separation of powers between the agencies responsible for policy design and those responsible for policy enforcement. In the US, farm policy originates in Congress (it must also be agreed to by the President) and is implemented by the US Department of Agriculture (Moyer and Josling, 1990; Gardner, 1987a). In the EU, the amount of intervention

¹ The basic analysis of agricultural policy, including quotas, is found in Gardner (1983, 1987a, 1987b). Analysis of the US peanut program, EU dairy policy, and Canadian supply management can be found in Rucker and Thurman (1990), Buckwell (1997), and Schmitz (1983), respectively.

² An economic analysis of cheating on different means of income redistribution (such as coupled and decoupled farm subsidies) can be found in Giannakas and Fulton (2000b, 2002). Cheating on farm subsidies translates into a misrepresentation (i.e., false reporting) of the farm characteristic on which payments are based and, unlike cheating on output quotas, it does not affect the actual production of the subsidized commodity.
is decided by the Council of Agricultural Ministers and the policy is implemented by the member countries of the EU, each of which differ significantly in the political weight they assign to their respective farm sectors (Swinbank, 1997; From and Stava, 1993). In Canada, agricultural policy legislation is introduced by the Cabinet to the House of Commons, which must pass it before it becomes law and is put into effect by Agriculture and Agri-Food Canada. Within this general framework, the year-to-year administration may be further devolved. For supply management, for instance, provincial quotas for chicken are set by a national agency, while the provincial marketing boards are responsible for enforcing these quotas.

To capture this separation of powers, an institutional arrangement characterized by decentralized policy making is adopted in this study. This richer policy making structure, which considers separately the decisions of the policy maker and the policy enforcer, is required because differences in the political preferences of the enforcement agency are shown to significantly affect the level of enforcement and the incidence of output quotas.3

Enforcement policy design is modeled in this paper as a sequential game between an enforcement agency that determines the level of enforcement of the quota program (designed by the regulator at an earlier stage in the game), and the farmers who make the production decisions. The payoff functions of the enforcement agency and the farmers are assumed to be common knowledge. The enforcement agency moves first and decides on the degree to which the quota will be enforced, knowing exactly how her choices will affect the producer optimizing decisions and welfare. Once the enforcement policy is determined, the producers decide on the quantity to produce while observing both the quota level and the enforcement policy in place. Different scenarios concerning the political preferences of the enforcement agency and the decision variables (i.e., enforcement parameters) it controls are examined within this framework.

To avoid Nash equilibria involving non-credible strategies, all formulations of the sequential game developed in the paper are solved using backwards induction (Gibbons, 1992). Thus, the paper is structured so that the problem of the farmers is considered first, followed by the solution to the enforcement agency’s problem. This latter outcome determines the subgame perfect equilibrium enforcement and production of the commodity. The welfare effects of enforcement costs and cheating are considered in the context of a static, partial equilibrium, closed economy model.

The title of the paper stems from a major result of the analysis, namely that quota enforcement increases with the weight placed by the enforcement agency on producer welfare. The paper shows that although violation of the quota limit might be optimal for the individual producer, producer welfare falls when over-production of the regulated commodity occurs. Thus, an enforcement agency operating with the interests of producers in mind will practice "tough love" – restricting the very actions that producers would prefer to undertake.

2. Output Quotas and Producer Behavior4

Output quotas have traditionally been a common means of market intervention by policy makers who desire to transfer income to agricultural producers. Supply restrictions result in surplus transfers from consumers to producers through their effect on market price while they do away with the need for policy makers to raise funds through taxation. Figure 1 depicts the traditional

---

3 Note that, by considering a centralized policy making structure, the analysis of GF implicitly assumes identical political preferences of the regulatory and enforcement agencies involved in agricultural policy making.

4 The analysis in this section is an expanded (detailed) version of the producer problem analyzed in GF and it is presented here for completeness of exposition.

Figure 1: Welfare Effects of Output Quotas.

Assuming that producers are the residual claimants of the returns generated from the sales of the product less the cost associated with its production, a quantity restriction at \( Q \) increases producer surplus by the area \( p(Q)ADp^e - DBC \) while consumer losses from the increased prevailing market price are given by the area \( p(Q)ABp^e \). The distortions in the use of productive resources under the program generate a deadweight welfare loss (DWL). The distortionary cost of market intervention is equal to the triangle \( ABC \). The policy mechanism has no effect on the welfare of taxpayers; income is redistributed directly from consumers to producers. The implicit assumption in this analysis is that producers comply completely with the provisions of the farm program, i.e., farmers do not cheat.

\[ p(Q^*) \]

---

5 This assumption covers the case when the producers and quota holders are the same person; it also covers the case when the producer leases the quota from the quota holder at a fixed quota rental rate.
As pointed out by GF, however, given the increased benefits from the quota program, producers may find it economically optimal to cheat by violating the quota limit, i.e., to over-produce and sell the above-quota amount through alternative channels. Assuming that producers know with certainty the (distorted) market price, the penalty charged per unit of over-produced and detected quantity, and the probability that they will be investigated, the decision of the individual producer as to whether he will over-produce (and if so, by how much) can be modeled as decision making under uncertainty.

Assuming producer neutrality toward risk, the producer has the choice between a certain outcome (i.e., his profits if he does not cheat) and the expected payoff in case of over-production. The expected benefits for the producer that violates the quota limit are given by the revenues from production, minus the cost of production and the expected penalty on above-quota quantity. More specifically, the problem faced by the representative risk-neutral producer of the regulated commodity can be written as:

\[
\max_{q_m} E[\pi] = p\left(\bar{Q} + Q_m\right)\bar{q} + q_m - c\left(\bar{q} + q_m\right) - \delta \rho q_m
\]

where \(q_m\) is the quantity produced by the representative farmer over and above his quota limit \(\bar{q}\) (the total quantity produced is \(q^* = \bar{q} + q_m\)); \(p(\bar{Q} + Q_m)\) is the market price when the industry quota is set at \(\bar{Q}\) and the aggregate amount of cheating by the \(N\) representative producers equals \(Q_m\); \(c(\bullet)\) is the cost function; \(\rho\) is the penalty per unit of over-produced and detected quantity; and \(\delta\) is the probability that the producer will be detected (and penalized) in case he cheats on the farm program. Implicit in the formulation of the problem presented in equation (1) is the assumption that the representative producer holds competitive conjectures; he does not perceive that he has any impact on aggregate output.

The detection probability takes values between zero and one (i.e., \(\delta \in [0, 1]\)), and is assumed to be a linear function of the quantity over-produced, i.e., \(\delta = \delta_0 + \delta_1 q_m\). This formulation of the detection probability captures the idea that the more a producer cheats, the greater is the likelihood that cheating will be detected (see GF). The intercept of the detection probability function, \(\delta_0\), reflects the probability that the farmer will be audited. The slope of the detection probability function (i.e., the change in \(\delta\) caused by a change in quantity over-produced), \(\delta_1\), is assumed strictly positive and exogenous to policy enforcers. The parameter \(\delta_1\) is assumed to depend on the observability of producers’ actions by third parties and

---

6 The model in equation (1) can be modified to include aversion toward risk by the representative producer and/or private costs from cheating. The risk averse farmer chooses \(q_m\) to maximize his expected utility (i.e., \(\max_{q_m} E[U(\pi)] = (1 - \delta) \mathbb{E}[q^* + q_m - c(q^* + q_m)] + \delta \mathbb{E}[q^* + q_m - c(q^* + q_m) - \rho q_m]\).

The aversion of the representative farmer toward risk results in reduced cheating relative to the risk neutral producer of the regulated commodity. Cheating is also reduced when potential costs incurred by producers in protecting themselves from detection (i.e., \(\mu q_m\)) are incorporated into the representative producer’s objective function. Even though both risk averse behavior and private costs from cheating change the results quantitatively, the qualitative nature of the results in this study remains unaffected.

7 In the context of this paper, audits are regarded as random; policy enforcers determine and announce the number of producers that will be investigated, \(\tilde{N}\), and the identities of the particular producers are determined at random. On this basis, for any particular producer, the probability of being audited, \(\delta_0\), is given by the percentage of the farmers that will be investigated, i.e., \(\delta_0 = \tilde{N}/N\).
the social attitudes towards cheating; the degree to which the third party that observes the illegal behavior will report it to policy enforcers. The very existence of a USDA “hotline” where cases of “fraud” can be reported (obviously by third parties) indicates that the variable component of the detection probability is considered an important element of policy enforcement (USDA Office of Inspector General, 2001).

Optimization of the representative producer’s objective function yields the following first order condition for a maximum:

\[ p(\bar{Q} + Q_m) = c(q + q_m) + (\delta_0 + 2\delta_1 q_m) \rho \]  

Equation (2) indicates that the farmer will produce up to the point where the market price (presumed by the producer to be exogenous), \( p(\bar{Q} + Q_m) \), equals the marginal cost (mc) of production, \( c(q + q_m) \), plus the marginal penalty \( (\delta_0 + 2\delta_1 q_m) \rho \). The marginal penalty, \( mp \), is the change in the expected penalty function for a change in the above-quota production, where the expected penalty function is \( pf = \delta_1 q_m \).

Aggregating over the \( N \) representative producers of the regulated commodity and assuming symmetry at equilibrium, the total over-quota production, \( Q_m = Nq_m \), is given by the expression:

\[ p(\bar{Q} + Q_m) = C(\bar{Q} + Q_m) + (\delta_0 + 2\delta_1 q_m) \rho \]  

where \( C(\bar{Q} + Q_m) \) is the industry supply curve and \( \delta_1 = \delta_1 / N \).

Figure 1 graphs equation (3). The equilibrium production, \( Q^* = \bar{Q} + Q_m \), is determined by the intersection of the (downward sloping) demand curve, \( D \), with the vertical summation of the industry supply curve, \( S \), and the \( MP \) curve. The \( MP \) curve is the horizontal summation of individual producers’ \( mp \) curves. The \( D \) and \( S \) curves are assumed linear while the linearity of the \( MP \) curve comes from the fact that the penalty function is quadratic in \( Q_m \). The \( MP \) curve has the same intercept as the average penalty function, \( APF = (\delta_0 + 2\delta_1 q_m) \rho \), when those are graphed relative to the origin of \( \bar{Q} \) in Figure 1, while the slope of the \( MP \) curve is twice the slope of the \( APF \). Although linearity of \( D \), \( S \), and \( MP \) has been employed for the purposes of this paper, more general formulations can be used without changing the qualitative nature of our results.

For linear approximations of the demand and supply curves (i.e., \( D(Q^*) = a_0 + a_1 Q^* \) and \( S(Q^*) = b_0 + b_1 Q^* \) where \( a_1 < 0 < b_1 \) and \( a_0 > b_0 \)), the aggregate above-quota production is:

\[ Q_m = \frac{(b_1 - a_1)(Q^e - \bar{Q}) - \delta_0 \rho}{b_1 - a_1 + 2\delta_1 \rho} \]  

where \( Q^e \) is the undistorted equilibrium production. All other variables are as previously defined.

Equation (4) indicates that the above-quota production depends on market conditions, the policy variable \( \bar{Q} \), and the detection probability and penalty parameters. With respect to the enforcement parameters, the analysis indicates that \( Q_m \) will be positive when \( \delta_0 < \frac{b_1 - a_1}{\rho} \) (or, alternatively, when \( \rho < \frac{b_1 - a_1}{\delta_0} \)). These critical values of \( \delta_0 \) and \( \rho \) are denoted as \( \delta_0^{nc} \) and \( \rho^{nc} \) respectively, where the superscript \( nc \) stands for
no cheating. Of course, values of the enforcement parameters in excess of \( \delta_0^{nc} \) or \( \rho^{nc} \) result in complete deterrence of cheating (i.e., \( Q_m = 0 \)).

Comparative static results can be seen by examining Figure 1. An increase in \( \delta_0 \) results in a parallel upwards shift of the \( S+MP \) curve in Figure 1 and reduced cheating. An increase in \( \rho \) increases both the intercept and the slope of the \( S+MP \) curve; the intercept shifts upwards while, at the same time, the curve rotates leftwards. The result is a lower over-produced quantity and hence a lower total output.

When the combination of policy variables and enforcement parameters is such that cheating occurs, the welfare effects of the policy instrument change. Figure 1 shows the welfare effects of cheating on output quotas. Relative to the “perfect and costless enforcement” situation examined in the traditional analysis of output quotas, cheating results in reduced producer surplus, increased consumer surplus and the introduction of taxpayers among the interest groups of the policy.

In particular, for above-quota production \( Q_m \), producers lose the area \( p(\overline{Q})AEp(\overline{Q}^*)-EFHI \) and consumers gain the area \( p(\overline{Q})AFp(\overline{Q}^*) \). Taxpayers gain the area \( CGHI \) that corresponds to penalty payments on detected above-quota quantity. There are also the resource costs of monitoring and enforcement that, though not present in Figure 1, have to be taken into account. These costs, denoted as \( \Phi(\delta_0) \), are assumed to be a non-decreasing function of \( \delta_0 \) (i.e., \( \Phi'(\delta_0) \geq 0 \), \( \Phi''(\delta_0) \geq 0 \)) and have to be included into both taxpayer costs\(^{10} \) and welfare losses from market intervention. Specifically, the change in taxpayer surplus in the presence of cheating equal \( (1+d)[CGHI-\Phi(\delta_0)] \) where \( d \) is the marginal deadweight loss from taxation (Fullerton, 1991; Ballard and Fullerton, 1992).\(^{11} \) Finally, the DWL of market intervention is decreased by area \( AFGC \) and increased by \( (1+d)\Phi(\delta_0) \) relative to the “costless deterrence” approach.

Before moving into the analysis of the optimal decisions of the agency responsible for the enforcement of the quota program, it is important to emphasize that, while above-quota production is the optimal decision for the individual producer that holds competitive conjectures (since this extra production is not perceived as affecting the market conditions), aggregate noncompliance increases total production and reduces price and total producer welfare. Note that this situation is similar in spirit to the well-known outcome of the Prisoners’ Dilemma. While there is no strategic interdependence between agricultural producers/quota holders in the case examined in this paper (i.e., farmers do not believe that their actions affect the payoff of other producers), the outcome of their optimizing behavior is a Pareto inferior situation characterized by welfare losses for producers as a whole.

---

8 Assuming that the per unit penalty exceeds the gains by producers in violating the quota limit at the margin (i.e., \( \rho > p(\overline{Q})-c'(\overline{Q}) = (b_1-a_1)(\overline{Q}^* - \overline{Q}) \)), implies that \( \delta_0^{nc} \) is less than one.

9 Recall that the welfare of producers is maximized at the monopoly output. Any increase of the supplied quantity above the monopoly output results in reduced producer surplus. Thus, above-quota production reduces producer surplus whenever the regulated quantity is set at, or more than, the monopoly output.

10 An alternative scenario could be the one where the enforcement costs are incurred by the producers of the regulated commodity. In such a case, \( \Phi(\delta_0) \) is subtracted from producer surplus and the taxpayers are the net beneficiaries of the program receiving the revenues from penalties (i.e., area \( CGHI \)).

11 It should be mentioned that there are also fixed budgetary costs arising from the operation of the agency responsible for agricultural policy enforcement. These costs are not associated with the existence of a specific farm program, however. Instead, the operation of the enforcement agency is due to government intervention in agriculture. Since the existence of the enforcement agency is not contingent upon the presence of any farm program in particular, the fixed costs from the operation of the enforcement agency are not incorporated into the taxpayer costs from output restrictions.
3. Optimal Enforcement by the Enforcement Agency

The analysis above shows that the above-quota production depends on the quota level $Q$ set by the regulator and the detection probability and penalty parameters $\delta_0$, $\delta'_1$, and $\rho$ (see equation (4)). The penalty $\rho$ on detected above-quota quantities is usually set elsewhere in the legal system and is therefore exogenous to agricultural policy enforcers, as is the detection probability parameter $\delta'_1$. With the enforcement parameters $\delta'_1$ and $\rho$ being exogenous, the only avenue policy enforcers have for influencing producer behavior is through the choice of the audit probability $\delta_0$.

This section of the paper examines the problem of the enforcement agency as it determines the optimal enforcement of an output quota $\bar{Q}$ set by the regulator, knowing exactly how its decisions will affect the behavior of the farmers and the welfare of all interest groups borne by the policy. In particular, the objective of the enforcement agency is the determination of the audit probability that maximizes its objective function. For linear approximations of the supply and demand curves, the problem of the enforcement agency can be written as:

$$
\max_{\delta_0} W = kCS + \theta PS + TS = \\
k \left\{- \frac{1}{2}a_1 \left[ Q + Q_m \right]^2 \right\} + \\
+ \theta \left\{ (b_h - a_i) \left[ Q + Q_m \right]Q^e + \left\{ a_1 - \frac{1}{2}b_i \left[ Q + Q_m \right]^2 - \left( \delta_0 + \delta'_1 Q_m \right) \rho Q_m \right\} + \\
+ (1 + d) \left( \delta_0 + \delta'_1 Q_m \right) \rho Q_m - \Phi(\delta_0) \right\}
$$

$$
(5)
$$

s.t. 

$$
Q_m = \frac{(b_h - a_i) \left(Q^e - \bar{Q}\right) - \delta_0 \rho}{b_1 - a_1 + 2\delta'_1 \rho} \\
\delta_0 \leq \frac{(b_h - a_i) \left(Q^e - \bar{Q}\right)}{\rho}
$$

where $CS$, $PS$ and $TS$ stand for consumer surplus, producer surplus and taxpayer surplus respectively. The parameters $\theta$ and $k$ represent the weights placed by the enforcement agency on producer surplus and consumer surplus respectively. All other variables are as previously defined.

The problem specified in equation (5) is a simple, static optimization problem with both equality and non-equality constraints. The equality constraint reflects the reaction function of the producers of the regulated commodity (equation (4)). The non-equality constraint requires that the optimal audit probability should not exceed the level that eliminates cheating. The reasoning is as follows. Since monitoring requires resources, the only effect of an increase in $\delta_0$ above $\delta_0^{ac}$ would be the growth of $\Phi(\delta_0)$. Alternatively, the inequality constraint can be seen as requiring $Q_m$ to be greater than, or equal to, zero.

---

12 The optimal enforcement of output quotas when penalties are endogenous to policy enforcers is considered in the next section of the paper.
Assuming enforcement costs of the form $\Phi(\delta_0) = \psi \delta_0^2 / 2$ (where $\psi$ is a strictly positive scalar depending on things like the agrarian structure and the number of producers), the Lagrangean of the enforcement agency’s problem is:

$$L = k \left\{- \frac{1}{2} a_1 [Q + Q_m] \right\} + \theta \left\{ (b_1 - a_1) [Q + Q_m] Q^e \right\} + \gamma \left\{ \left( 1 - \frac{1}{2} b_1 \right) [Q + Q_m] \right\} -$$

$$- \left\{ \theta - 1 - d \right\} (\delta_0 + \delta_1 Q_m) \rho Q_m - (1 + d) \psi \delta_0^2 + \lambda \left\{ \left( \frac{b_1 - a_1}{\rho} (Q^e - \bar{Q}) - \delta_0 \right) \right\}$$

while the Kuhn-Tucker conditions for a maximum are:

$$L_{\delta_0} = a_1 (b_1 - a_1) \rho Q^e + (1 + d) (b_1 - a_1) [Q - \bar{Q}] \rho - 2 a_1 \delta_1 \rho^2 (\theta - k) + \gamma \left\{ a_1 [Q - \bar{Q} \right\} [b_1 + 2 \delta_1 \rho] Q^e \rho^2 - \frac{(b_1 - a_1) \rho [Q - \bar{Q}] \rho^2}{\left(b_1 - a_1 + 2 \delta_1 \rho \right)^2} \delta_0 - \lambda \leq 0, \quad \delta_0 \geq 0 \rightarrow L_{\delta_0} \delta_0 = 0$$

$$L_\lambda = \frac{(b_1 - a_1) (Q^e - \bar{Q})}{\rho} - \delta_0 \geq 0, \quad \lambda \geq 0 \rightarrow L_\lambda \lambda = 0$$

Solving the optimality conditions for $\delta_0$ it can be shown that, for given market conditions and resource costs of monitoring, the optimal audit probability depends on the relative weight placed by policy enforcers on the welfare of the producers of the commodity. More specifically, the K-T conditions imply that whenever $\theta$ is less than a critical value $\theta_c$, where

$$\theta_c = \frac{(1 + d) (b_1 - a_1) \rho [Q^e - \bar{Q}] + \gamma \left\{ a_1 \rho^2 (b_1 + 2 \delta_1 \rho) \right\}}{-a_1 [Q^e - \bar{Q}] \rho} + k$$

$\delta_0$ will equal:

$$\delta_0^{\theta-c} = \frac{a_1 (b_1 - a_1) \rho Q^e + (1 + d) (b_1 - a_1) [Q - \bar{Q}] \rho - 2 a_1 \delta_1 \rho^2 (\theta - k) + \gamma \left\{ a_1 [Q - \bar{Q}] \right\} [b_1 + 2 \delta_1 \rho] Q^e \rho^2}{(1 + d) \left\{ \psi [b_1 - a_1 + 2 \delta_1 \rho]^2 + 2 \rho^2 [b_1 - a_1 + \delta_1 \rho] \right\} + a_1 \rho^2 - \left(b_1 + 2 \delta_1 \rho \right) \rho^2}$$

13 When the enforcement agency audits $\tilde{N}$ producers the cost of program enforcement with audit probability $\tilde{\delta}_0$ can be written as $\Phi(\tilde{\delta}_0) = \tilde{N} \phi(\delta_0)$ where $\phi(\delta_0)$ is the cost incurred by the enforcement agency per producer. The cost per producer when the agency enforces an audit probability of $\delta_0$ can be expressed as $\phi(\delta_0) = \zeta \delta_0$, where $\zeta$ depends on things like the agrarian structure, the dispersion of the farms, etc. The total enforcement costs can then be written as $\Phi(\delta_0) = N \zeta \delta_0^2$. Setting $N \zeta = \psi / 2$ we get the expression of the enforcement costs used in this paper.
If $\theta$ is greater than or equal to $\theta_c$, cheating will be completely deterred by an audit probability that equals $\delta_0^{nc}$, i.e.,

$$
\delta_0^{\theta \geq \theta_c} = \frac{(b_1 - a_1)(Q^e - \bar{Q})}{\rho}
$$

where the superscripts denote the relative weight placed by the enforcement agency on the welfare of producers of the regulated commodity.

The optimality conditions indicate that an increase in $\theta$ increases the optimal enforcement (i.e., $\partial \delta_0^{\theta=\theta_c} / \partial \theta > 0$) and the likelihood that cheating will be completely deterred (i.e., the likelihood that $\theta \geq \theta_c$). Since $\theta_c$ is an increasing function of $\psi$ and $k$, the likelihood that cheating will be completely deterred falls with an increase in either the resource costs of monitoring and/or the weight placed on consumer welfare. The reasoning is as follows. In the previous section it has been shown that the greater is the above-quota production, the lower is the producer surplus and the greater is the consumer surplus. Therefore, increased significance of producers dictates increased enforcement and reduced cheating while increased significance of consumers dictates the opposite.¹⁴

Based on the previous results it is easy to determine the optimal enforcement in the limiting case where the enforcement agency places no weight on producers and consumers of the regulated commodity but its objective instead is to minimize taxpayer costs from cheating.¹⁵ In such a case, the level of monitoring is derived by substituting zero for $\theta$ and $k$ into equation (7) i.e.,

$$
\delta_0^{\theta=k=0} = \frac{(b_1 - a_1)^2(Q^e - \bar{Q})}{\psi[b_1 - a_1 + 2\delta_1^0 \rho] + 2\rho^2[b_1 - a_1 + \delta_1^0 \rho]}
$$

where the superscript denotes the weight attached by the enforcement agency to producers and consumers.

Equation (9) indicates that when $\theta=k=0$ the optimal enforcement is always positive while, comparing $\delta_0^{\theta=k=0}$ to $\delta_0^{nc}$, it can be shown that $\delta_0^{\theta=k=0}$ will always be less than the $\delta_0$ that completely deters cheating. Thus, the optimal policy under $\theta=k=0$ will always involve allowance of some cheating. Since the weight placed by policy enforcers on producer surplus is the lowest possible, the level of enforcement is minimized. Obviously, taxpayers incur a loss whenever the

---

¹⁴ An alternative interpretation of the optimality conditions is as follows. For given market conditions and weights on the interest groups ($\theta$ and $k$), the optimal $\delta_0$ will be given by equation (7) whenever the resource costs of monitoring ($\psi$) are greater than a critical value $\psi_c$, where $\psi_c = \left[ - (1 + d)(b_1 - a_1)(Q^e - \bar{Q})\rho^2 - a_1(\theta - k)\rho^2\bar{Q} \right] / \left[ (1 + d)(b_1 - a_1)(b_1 - a_1 + 2\delta_1^0 \rho) \right]$. For relatively low enforcement costs (i.e., $\psi \leq \psi_c$) the optimal response of the enforcement agency will be the complete deterrence of cheating (i.e., $\delta_0 = \delta_0^{nc}$).

¹⁵ The payoff function of the enforcement agency when $\theta=k=0$ is measured by the regulator’s revenues net of monitoring costs. Alternatively, the enforcement agency can be seen as seeking the $\delta_0$ that minimizes the total budgetary costs from cheating, i.e., the resource costs of investigation minus the penalties collected from producers detected cheating.
optimal $\delta_0$ is greater than $\delta_0^{\theta=k=0}$, i.e., the enforcement costs exceed the budgetary benefits from enforcement.

Manipulating $L_{\delta_0}$ it can be shown that when $\theta \in [0, \theta_c)$, the optimal monitoring is determined by equating the marginal benefits from enforcement ($MB_e$) with the marginal costs of enforcement ($MC_e$), i.e.,

$$a_1(b_1-a_1)k\rho Q^e + (1+d)(b_1-a_1)^2(Q^e-Q)\rho - 2a_1\delta^2 \bar{Q}(\theta-k) + (b_1-a_1)[-a_1Q-(b_1+2\delta\rho)Q^e-Q]\rho \delta_0 = (1+d)k\delta_0$$

where the terms on the LHS and the RHS of the equation sign are the $MB_e$ and the $MC_e$ respectively. The marginal benefits from enforcement include penalties collected on detected above-quota quantities and the benefits from induced honesty. These latter benefits include the consequences for interest group welfare of increased enforcement and reduced above-quota production. Graphically, the $MB_e$ curve can be seen as a linear downward sloping straight line in the relevant area for program enforcement (i.e., $0 \leq \delta_0 \leq \delta_0^{nc}$). The $MB_e$ curve is downward sloping due to the decrease in cheating caused by increases in $\delta_0$.

The position of the $MB_e$ curve depends on the relative weights placed by the enforcement agency on the well being of the interest groups, while the resource costs of monitoring determine the position (slope) of the $MC_e$ curve. When the enforcement agency places no weight on producers and consumers, the optimal $\delta_0$ is determined by the intersection of the $MC_e$ curve with the $MB_e^{\theta=k=0}$ curve (Figure 2, Panel (a)).

An increase in the relative value of $\theta$ will shift the $MB_e$ curve upwards, increasing both its intercept and its slope. For a positive, but relatively low $\theta$ (i.e., $\theta \in (0, \theta_c)$), $\delta_0$ is determined by the intersection of the $MB_e^{\theta=\theta}$ curve and the $MC_e$ curve in Panel (a) of Figure 2. When $\theta$ is relatively high (i.e., $\theta > \theta_c$), the optimal $\delta_0$ will equal $\delta_0^{nc}$ since, in this case, the $MB_e$ and $MC_e$ curves do not intersect inside the relevant region for policy enforcement (i.e., the area to the left of $\delta_0^{nc}$). Note that the $MC_e$ curve meets the $MB_e$ curve at $\delta_0^{nc}$ whenever $\theta = \theta_c$. Obviously, for any given $\theta \in [0, \theta_c)$, the greater are the resource costs of monitoring, the greater is the slope of the $MC_e$ curve, and the lower is the level of enforcement.

---

16 Note that when $\theta < \theta_c$, $0 < \delta_0 < \delta_0^{nc}$. From K-T conditions $L_{\delta_0}$ and $\lambda$ will equal to zero.

17 The reverse is true for an increase in the weight placed on consumers. An increase in $k$ will shift the $MB_e$ downwards by decreasing its intercept and slope resulting in lower $\theta_c$ and, thus, reduced likelihood that cheating will be deterred.

18 Notice that while Panel (a) of Figure 2 illustrates the case of increasing marginal enforcement costs, the marginal costs from enforcement can in fact be constant. In such a case, the relevant $MC_e$ curve is a horizontal line that meets the vertical axis in Panel (a) of Figure 2 at the level of the constant marginal costs.
Figure 2: Optimal Enforcement and Strategic Interdependence between the Enforcement Agency and the Farmers.

Figure 2 also graphs the strategic interdependence between the enforcement agency and the producers; it shows the effect enforcement decisions have on cheating. Panel (b) depicts the cheating equilibrium for the \( N \) representative producers. The above-quota production \( Q_m \) is determined by the intersection of the industry demand curve with the relevant \( S+MP \) curve i.e., the \( S+MP^{\theta > \theta_c} \), \( S+MP^{\theta < \theta_c} \), or the \( S+MP^{\theta = k=0} \) curve depending on the political preferences of policy enforcers. An increase in \( \delta_0 \) causes a parallel upward shift of the \( S+MP \) curve which results in reduced above-quota production.
Mathematically, the equilibrium \( Q_m \) under the alternative scenarios concerning the weights placed by policy enforcers on the welfare of the interest groups can be derived by substituting the relevant \( \delta_0 \) into the producers’ reaction function in equation (4). Since enforcement increases with an increase in the relative weight placed by policy enforcers on producer surplus, the greater is \( \theta \), the lower is the above-quota production (i.e., \( Q_m^{\theta=k=0} > Q_m^{\theta<\theta_c} > Q_m^{\theta=\theta_c} = 0 \)).

Overall, for any given output quota, the greatest monitoring and the lowest over-production will occur when producer surplus is weighted highly and consumer surplus has no weight for program enforcers. On the other hand, enforcement is minimized and cheating is maximized when the enforcement agency places no weight on producers and consumers.\(^{19}\)

4. Extension of the Model - Endogenous Penalties

The previous analysis and results are based on the assumption that penalties are exogenous to agricultural policy enforcers. This is generally the case since, as pointed out previously, penalties are typically set by the legal system. However, to complete the analysis of the optimal enforcement of output quotas, it is useful to consider the case where policy enforcers have control over penalties as well as the audit probability.

The main implication of endogenizing penalties is that, if there are no economic costs associated with setting \( \rho \), policy enforcement becomes costless. In particular, since both \( \delta_0 \) and \( \rho \) affect farmers’ production decisions, the enforcement agency could achieve its objectives by substituting costly monitoring with costless penalties.

To show this result, note that endogenizing penalties requires another K-T condition to the optimization problem of the enforcement agency presented by equation (5), i.e.,

\[
L_{\rho} \leq 0, \quad \rho \geq 0 \quad \rightarrow \quad L_{\rho} \rho = 0
\]

Obviously, the optimal choice of an enforcement agency that places a relatively high weight on producer surplus would still be the complete deterrence of cheating. In this case however, the induced producer compliance will be achieved by the establishment of a very tiny probability of a very severe penalty on above-quota quantities. More specifically, whenever \( \theta \) is greater than or equal to a critical value \( \theta_c^\rho \), where

\[
\theta_c^\rho = \frac{(1 + d)(b_1 - a_1)(Q^e - \bar{Q}) - 2a_1k\bar{Q}}{(b_1 - 3a_1)\bar{Q} - (b_1 - a_1)Q^e}
\] (10)

the solution to the enforcement agency’s problem is:

\[
\delta_0^{\theta \geq \theta_c^\rho} \rho^{\theta \geq \theta_c^\rho} = \infty \quad (\text{with} \quad \text{Limit} \quad \delta_0^{\theta \geq \theta_c^\rho} \rho^{\theta \geq \theta_c^\rho} = 0)
\] (11)

\(^{19}\) The above result holds whenever the weight placed on producer welfare, \( \theta \), is greater than or equal to the weight placed on consumers, \( k \). In a situation where \( k > \theta \), the minimum enforcement and the maximum production will occur when the enforcement agency places zero weight on producers and positive weight on consumers. This is not however, a realistic assumption for a developed country’s agricultural policy makers.
When, however, \( \theta \) is less than \( \theta^\rho \), program enforcement is imperfect and some cheating occurs. The optimal enforcement consists of (almost) zero \( \delta_0 \) and a positive but smaller than \( \rho^{nc} \) penalty, i.e.,

\[
\delta_0^{\theta<\theta^\rho} \equiv 0 \quad \text{and} \quad \rho^{\theta<\theta^\rho} = \frac{(b_1-a_1)
\left[-(1+d)(b_1-a_1)\left(Q^e-\bar{Q}\right) - 2a_1k(Q^e-\bar{Q})Q^e - 2d(b_1-\bar{Q}-a_1)Q^e + \partial(b_1-a_1)\left(Q^e-\bar{Q}\right)^2\right]
}{2\delta\left[Q^e-\bar{Q}\right] - 2a_1(\theta-k)\bar{Q} - (\theta+1+d)(b_1-a_1)\left(Q^e-\bar{Q}\right)^2}
\] (12)

Consistent with \textit{a priori} expectations, an increase in the weight placed by program enforcers on producer welfare increases both the program enforcement and the likelihood that cheating will be completely deterred (i.e., the likelihood that \( \theta \geq \theta^\rho \)). Since the enforcement of output quotas will be perfect and costless when \( \theta \geq \theta^\rho \), one interpretation of the assumption of “perfect and costless policy enforcement” that is implicit in the traditional analysis of output quotas is that enormous fines can be costlessly levied on producers that violate their quota limit.\(^{20}\)

5. Concluding Remarks

Quotas have been commonly used by governments as a means of redistributing income in the economy. The increased price associated with supply restrictions creates an incentive for producers of the regulated commodity to cheat on the program by producing over and above the quota limit. Despite the incentives for, and the incidence of cheating on output quotas, the traditional analysis of the policy takes place under the assumption of perfect and costless enforcement.

This paper develops a sequential game theoretic model of decentralized policy making to examine the optimal enforcement and incidence of output quotas in the presence of enforcement costs and producer noncompliance with the provisions of the policy mechanism. While the focus in the paper is on agriculture, an area in which quotas are extensively used, the analysis applies to other situations where quotas are used, such as fisheries.

Analytical results show that, while violation of the quota limit might be optimal for the individual producer that holds competitive conjectures, above-quota production depresses the market price and results in welfare losses for the producers and gains for the consumers of the regulated commodity. The optimal level of enforcement (and, thus, cheating) depends on the relative weights placed by policy enforcers on the welfare of the interest groups, and the resource costs of monitoring producer compliance with the terms of the quota program. The greater is the weight placed by the enforcement agency on producer welfare and/or the lower are the monitoring costs, the greater is the level of enforcement of the output quota scheme. For given enforcement costs, enforcement of quotas is maximized when the enforcement agency places a high weight on producer welfare.

In addition to analyzing the optimal enforcement and incidence of output quotas, this paper highlights the conditions under which cheating is likely to be an issue. The section on Endogenous Penalties shows that if it is possible to costlessly levy enormous fines, then cheating will be effectively deterred in those situations where policy enforcers place a relatively high

\(^{20}\) Graphically, infinite per unit penalty results in an infinite slope of the \( S+MP \) curve faced by the producers of the regulated commodity.
weight on the surplus of producers. In short, the ability to levy very large fines essentially means that quota enforcement can be made both perfect and costless.

This result, however, raises the question as to whether large fines for quota limit violations are reasonable. The literature on the economics of crime provides some guidance and evidence on this issue. Specifically, it has been argued that severe punishment for minor law violations (i.e., Becker’s (1968) “optimal fine” result) is neither costless, credible nor just. The imposition of disproportionate fines would likely offend the public sense of justice; in short, justice requires that the punishment fit the crime (Carr-Hill and Stern, 1977; Stern, 1978; Stigler, 1970. On this point see also Shavell (1987), Kemp and Ng (1979), Polinsky and Shavell (1979), Landes and Posner (1975), Singh (1973), Tullock (1971), and Cowell (1990)).

If induced compliance through the establishment of enormous and costless fines is indeed infeasible, then cheating on output quotas will always be an issue and should be incorporated into economic analysis. This paper shows that cheating has important effects on the incidence of the policy, effects that vary with the political preferences of policy enforcers and the size of enforcement costs.

In addition to providing an understanding of the causes and consequences of cheating on output quotas, the results of this study can assist in explaining potential differences in compliance with policy rules observed in different areas/countries. Differences in the structure of the agricultural sector and the efficiency of institutions could account for differences in enforcement costs. Obviously, the greater is the number of producers of the regulated commodity and the more dispersed are the farms, the greater are the monitoring costs (the parameter $\psi$ in the models in this study). Increased monitoring costs mean less auditing and more cheating. Moreover, the smaller is the proportion of farm population, and/or the greater is the (perceived) difference between farm and non-farm incomes, and/or the more effective is the farm lobby, the more important politically the sector is expected to be (Becker, 1983; Gardner, 1987b). Increased weight on producers (i.e., a larger $\theta$) is translated into more enforcement and less cheating.

The pattern of excess quota production outlined in the examples in the Introduction is consistent with these results. The relatively low amount of above-quota production in the Canadian chicken industry can be interpreted in part as a consequence of enforcement having been devolved to the provincial market boards – since producers operate these boards, the weight attached to producer welfare can be expected to be high. As well, the cost of monitoring is expected to be relatively low, given the small number of chicken producers in any given province and the ease with which over-production can be detected. In the Italian milk case, the weight attached to producers can also be expected to be high, since the Italian government has a history of providing strong support to its agricultural producers. However, the large number of dairy producers in that country likely means that the cost of monitoring is high. As was shown, high monitoring costs translate into reduced enforcement, all other things the same.

There are limitations in the current study. As was posed at the outset, the focus of this study has been on the economic causes and consequences of farmer noncompliant behavior. Morality and culture, though significant determinants of individual behavior, are not incorporated into this analysis. For instance, there are many producers whose disutility from cheating would outweigh any expected benefit from violating the program rules. Simply put, there are people that would never cheat (for a discussion of the role of social conscience as a general deterrent to crime see Grasmick and Green (1980)). By focusing on the representative producer, the effects of personal attributes on cheating are not considered. Extensions of the model to account for producer heterogeneity and culture could provide a better understanding of the causes of cheating and

21 For instance, there are 2,815 chicken producers and 135 poultry processing plants (many of these are owned by one of the largest six processors) in all of Canada in 2002 (Chicken Farmers of Canada).
assist in further explaining discrepancies in terms of policy compliance observed between different areas/countries.

Finally, an interesting extension of this line of research involves the analysis of the optimal policy intervention, income redistribution, and transfer efficiency of output quotas in the decentralized policy making environment considered in this paper. The study of the optimal regulatory responses to enforcement imperfections and the efficiency of the policy mechanism in redistributing income in the economy are the subject of a companion article that is forthcoming in the next issue of this journal.

Acknowledgements

The authors wish to thank three anonymous Journal reviewers for useful comments on earlier versions of this paper. The usual caveats apply. Journal Series No. 13253, Agricultural Research Division, University of Nebraska-Lincoln.

Konstantinos Giannakas is Assistant Professor in the Department of Agricultural Economics at the University of Nebraska-Lincoln. He received his Ph.D. in Agricultural Economics from the University of Saskatchewan. His research interests include the areas of industrial organization of the agri-food marketing system, regulatory economics and agricultural policy, applied production economics, and the economics of agricultural biotechnology.

Murray Fulton is Professor of Agricultural Economics at the University of Saskatchewan. He received his Ph.D. from the University of California, Berkeley. His research and teaching interests are focused on industrial organization, agricultural industry analysis, co-operative theory, intellectual property and biotechnology.
References

---------------, “The Economics of Coupled Farm Subsidies under Costly and Imperfect Enforcement,” Agricultural Economics 22 (2000b), 75-90.

http://www.bepress.com/jafio/vol1/iss1/art2


