Cap-and-Trade Programs under Delayed Compliance: Consequences of Interim Injections of Permits*

Makoto Hasegawa† Stephen Salant‡

July 7, 2014

Abstract

Virtually every analysis of cap-and-trade programs assumes that firms must surrender permits as they pollute. However, no program, existing or proposed, requires such continual compliance. Some (e.g. the Acid Rain Program limiting SO_2 emissions) require compliance once a year; others (e.g. the Regional Greenhouse Gas Initiative limiting CO_2 emissions) require compliance every three years. The paths of emissions and permit prices would be invariant to compliance timing (Holland-Moore, 2013) if the government never injected additional permits between successive compliance dates. However, virtually all emissions trading programs require such injections through either (1) interim permit auctions or (2) sales from "cost containment reserves" intended to cap permit prices. In such cases, analyses which abstract from delayed compliance may mislead policy makers. For example, a cost containment reserve judged sufficient to cap prices at a ceiling over a year may sell out in a single day.

*We thank Andrew Stocking for alerting us to the delayed compliance provision and Soren Anderson for circulating an earlier draft of our paper to Stephen Holland and others concerned with AB-32. Severin Borenstein, Dallas Burtraw, Christopher Costello, Harrison Fell, Charles Kolstad, Gary Libecap, Carol McAusland, Michael Moore, and Matthew Zaragoza-Watkins provided helpful comments on an earlier draft. We thank the co-editor, Lucas Davis, and the anonymous referees of this journal for constructive suggestions that have greatly improved the paper. Hasegawa acknowledges the financial support provided by the Nakajima Foundation and the Yamada Academic Research Fund. Salant completed this research while in residence at Santa Barbara's Bren School as a UCE3 Senior Fellow.

†National Graduate Institute for Policy Studies (GRIPS). E-mail: m-hasegawa@grips.ac.jp.

‡Department of Economics, University of Michigan and Resources for the Future. E-mail: ssalant@umich.edu.

The Cost Containment Reserve (CCR) was used for the first time in [the March 5, 2014 RGGI] Auction 23. The demand for CO_2 allowances from bids submitted above the CCR Trigger Price of $4.00 exceeded the Initial Offering of 18,491,350 allowances and was sufficient to purchase all 5 million 2014 CCR allowances. After the CCR was exhausted, the auction cleared at a price of $4.00 per ton. There are no other CCR allowances available for sale in 2014.—Market Monitor Report (March 7, 2014)

1 Introduction

Cap-and-trade programs have been used in the past to solve the acid rain problem in the U.S. and are now being utilized at home and abroad to combat global warming. Such regulations may differ in their details. Nonetheless they share two important features.

First, although firms subject to the regulations are required to surrender permits to cover their emissions, they are never required to surrender permits on a continual basis ("continual compliance") but only at compliance dates (or "true ups") which recur periodically. As a result, a regulated firm may emit sulfur dioxide or carbon dioxide without possessing the permits to cover its emissions as long as it acquires the requisite permits by the compliance date. We refer to this aspect of the regulations as "delayed compliance." For example, under the U.S. Acid Rain Program (ARP) or the European Union Emission Trading Scheme (EU-ETS), compliance is required only once a year. In the case of the three federal bills to regulate carbon emissions in the U.S., none of which became law, compliance would have been required only once a year.[1] Other programs have even longer intervals between true-ups. For example, the Regional Greenhouse Gas Initiative (RGGI) has a compliance period of three years and so does California's AB-32 (although in the latter case, a fraction of the permits must be surrendered earlier as a down-payment).[2]

These programs typically share a second feature: they mandate injections of additional permits between true-up dates. Even the older programs have auctions between compliance

[1] Waxman-Markey's "American Clean Energy and Security Act of 2009," Kerry-Boxer's "Clean Energy Jobs and American Power Act of 2009," and Kerry-Lieberman's "American Power Act of 2010."

[2] In addition, as Holland and Moore (2014, p. 673) note, the Western Climate Initiative (2010) and Midwestern Greenhouse Accord (2010) define three-year compliance periods in their draft model rules.

dates. For example, ARP holds an auction of additional permits one month after the annual true-up while EU-ETS holds weekly auctions within each compliance period. California's AB-32 has an initial allocation of permits supplemented by subsequent injections of additional permits during the compliance period, and a similar plan was proposed in the three Congressional bills which died in Congress. All four of these programs prescribe a periodic sequence of auctions.

Recently, another method of injecting additional permits within the compliance period has become popular—the sale of additional permits from the regulator's stockpile. The intention is to have a reserve large enough to keep permit prices from soaring above a designated ceiling but not so large that it would compromise the pollution cap. RGGI has its "cost containment reserve" (CCR) and AB-32 has its "Allowance Price Containment Reserve" (APCR) for this purpose. The Kerry-Lieberman (2010) bill proposed sales of permits at a fixed price over a designated time interval terminating at the end of the interval—or sooner if the cost containment reserve was exhausted. Although EU-ETS and ARP currently lack such price collars, each program has considered using them. Some scholars have argued that the absence of such a "safety valve" in the EU-ETS constitutes a serious "design flaw."[3]

All of these programs allow storage of permits ("banking") for later use and most of these programs prohibit or severely restrict the opportunity to borrow from future allocations. For the reader's convenience, we summarize the features of various cap-and-trade programs in Table 1.[4]

In fashioning these programs, policy makers often ask economists to predict the consequences of different designs. What are the consequences of different ceiling prices and the reserve sizes to defend them? What are the consequences of different auction frequencies, amounts, and minimum bids (reserve prices)?

In response, a large literature has developed analyzing safety valves.[5] Burtraw et al.

[3]Stavins (2012) regards the absence of a safety valve or price collar in the European system as a "design flaw."

[4]This table amplifies information in Table 1 of Holland and Moore (2013).

[5]For a valuable explanation of the origins of the safety-valve concept and its evolution in the climate

Table 1: Summary Features of Cap-and-Trade Programs

Program	Compliance Timing	Fixed Price Sales	Interim Auctions	Permit Banking	Permit Borrowing
Acid Rain Program[2]	Annual	No	Yes, annual[1]	Unlimited	No
EU ETS[3]	Annual	No	Yes, weekly	Banking is allowed within each phase	Unlimited borrowing from next year's vintage of permits
Waxman-Markey	Annual	No	Yes, quarterly with reserve prices[4]	Unlimited	Borrowing from the next year's vintage of permits; Borrowing with interest permits with a vintage year 2-5 years later
Kerry-Lieberman[5]	Annual	Permits in "cost containment reserve" are sold at the fixed price within 90 days of true-up[6]	Yes, quarterly with reserve prices	Unlimited	Borrowing from the next year's vintage of permits; Borrowing with interest permits with a vintage year 2-5 years later
RGGI[7]	3-year period	Yes, sales from cost containment reserve at trigger prices to supplement auctions	Yes, quarterly with reserve prices[8]	Unlimited	Unlimited borrowing within 3-year compliance period
California AB-32[9]	3-year period with 30% annual down payment	Permits in allowance price containment reserve are divided into three equal-sized tiers and sold at fixed prices.[10]	Yes, quarterly with reserve prices[11]	Unlimited	Unlimited borrowing within 3-year compliance period

1. Additional permits are auctioned one month after annual true-up
2. SO_2 market established by Title IV of the 1990 Clean Air Act in the United States.
3. European Union Emission Trading Scheme.
4. 1-3% of permits are placed each year in "strategic reserve account" and sold in quarterly auctions with a reserve price.
5. The Kerry-Boxer bill has similar cost-containment provisions to the Waxman-Markey bill.
6. The fixed price is $25 per ton of CO2 equivalent in 2013 and increases in real terms by 5% each year thereafter.
7. The Regional Greenhouse Gas Initiative
8. Allowances allocated to the cost containment reserve are auctioned if auction prices exceed the trigger price. The trigger price is $4, $6, $8, and $10 per ton in each of 2014-17 years, respectively, and increase by 2.5% each year thereafter. The price floor is implemented by quarterly auctions with minimum reserve price. The reserve price is $2 per metric ton of CO2e in 2013 and increases by 2.5% each year thereafter.
9. California Assembly Bill 32, the Global Warming Solutions Act of 2006
10. In 2013 the three-tier fixed prices are $40, $45, and $50 per allowance and increase by 5% plus the rate of inflation each year thereafter.
11. The reserve price shall be $10 per metric ton of CO2 in 2013 and increase by 5 % plus the rate of inflation each year thereafter.

(2010) find that a price collar (also called a "symmetric safety valve" in the paper) outperforms a safety valve in a static setting. Fell and Morgenstern (2010) and Fell et al. (2012a) simulate a dynamic stochastic model of a cap-and-trade program with a price collar or a safety valve.[6] Fell and Morgenstern (2010) find that price collar mechanisms are more cost-effective than both purely quantity-based mechanisms and safety-valve mechanisms for a given level of expected cumulative emissions. They also find that the combination of a price collar with banking and borrowing systems can achieve expected cost as low as a tax with lower emissions variance. Fell et al. (2012a) find that *hard* collars, which ensure unlimited supply of reserve allowances to defend a ceiling price yield lower net present value of expected abatement costs than *soft* collars, price collars with limited supply of reserve allowances, for the same level of the expected cumulative emissions net of offsets. Recently, Hasegawa and Salant (2012) have shown that the price path of permits will remain constant as long as the government sells additional permits at its ceiling price and may collapse in response to government auctions even if they are anticipated. Remarkably, this entire literature (including our own contribution) has neglected to take account of compliance timing.[7]

The purpose of our paper is to explain why policy analyses assuming continual compliance may mislead policy makers if the delayed-compliance program analyzed also mandates interim injections of permits. To be constructive, we then provide a methodology for analyzing cap-and-trade programs under delayed compliance.

Before we begin, we would like to discuss briefly contributions to the literature that encourage the mistaken belief that compliance timing is an irrelevant detail. Holland and Moore (2013) is the only paper besides our own working papers to distinguish delayed from

context, see Jacoby and Ellerman (2004).

[6]In a dynamic context, intertemporal trading of emissions permits matters for economic efficiency. Cronshaw and Kruse (1996) and Rubin (1996) show that emissions trading allowing banking and borrowing of emission permits achieves the least-cost outcome. Like all dynamic analyses of cap-and-trade programs, both of these articles assume that firms are continually in compliance as well.

[7]Stocking (2012) analyzes the strategic actions of regulated firms in the presence of price controls in emissions permit markets and shows that, in attempt to reduce the equilibrium price of permits, firms may have incentive to purchase permits from the government at the ceiling price even when the prevailing market price of permits is lower.

continual compliance (or, as they call it, "prompt" compliance). Their paper provides a valuable table describing key institutional features of various cap-and-trade programs. It also contains a condition sufficient for the path of permit prices and emissions to be the *same* under continual and delayed compliance. As long as their invariance condition holds, analysts who assume continual compliance when analyzing cap-and-trade programs will not introduce error even though all such programs require only periodic true-ups. However, as Holland and Moore themselves warn "this sufficient condition might not hold, for example, if the regulator imposed price regulation or if the regulator injected or removed permits into the market" (Holland-Moore, p. 679). Elsewhere, they again note the limited applicability of their theorem: "Many existing and proposed programs include such price control mechanisms" (Holland-Moore, p. 673). Finally, in their very brief section 3.2 drawing on our prior work, they illustrate cases where the time paths of emissions and permit prices *are* sensitive to compliance timing.

Many of the simulation models in the literature (Fell and Morgenstern, 2010; Fell et al., 2012a; Fell et al., 2012b) also leave the impression that compliance timing is unimportant. Such models typically assume discrete time and define the period length in a way that obscures the distinction between delayed compliance and continual compliance. To understand this distinction, consider a discrete-time model where one period represents one day. If the government injects additional permits on some of the days within the next year, the policy of requiring that permits be surrendered every day to match that day's emissions (continual compliance in discrete time) *differs* from the policy of requiring permits to be surrendered once every 365 days to cover cumulative emissions during the entire year. If true-ups were required once every 365 days, injections of additional permits would occur *within* the compliance period. However, most of the discrete-time literature defines the length of each period to be the *same* as the length of the compliance period.[8] Since nothing by definition can

[8]There is an interesting parallel in the literature on storage of grains. In Samuelson (1957), grain harvests of the same size occur once every year and intraseasonal demand is stationary and deterministic. The price following every harvest begins low, grows to induce intraseasonal storage despite storage costs and foregone interest and collapses when the next harvest arrives, deterring interseasonal storage. Then the cycle repeats,

happen *between* periods, this seemingly innocuous modeling choice implicitly prevents examination of the consequences of a government injection of permits between one compliance date and the next.

To consider the effects of government policies conducted *within* a compliance period, we adopt a continuous-time formulation as less cumbersome than its discrete-time counterpart.

We proceed as follows. In section 2, we present a stark illustration of the errors that result if an analyst mistakenly assumes that firms are in continual compliance between true-up dates. Having established the need to replace this erroneous assumption with an appropriate one, we develop a straightforward methodology for doing so in section 3. Section 4 illustrates the application of this new methodology and shows that it can still be applied even when the regulations are complex. Section 5 concludes.

2 Why Compliance Timing Matters: a Preliminary Example

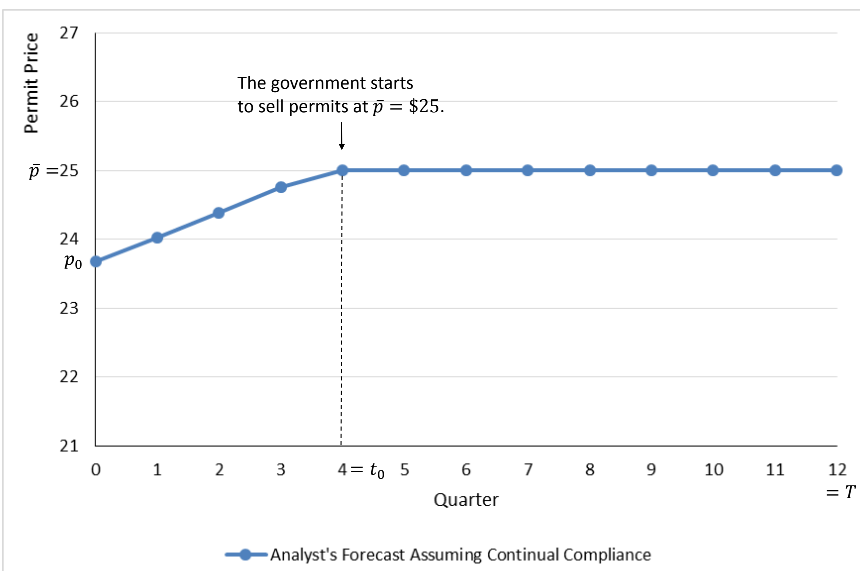
A common design is for the regulator to auction or grandfather $\bar{g}$ permits at the outset of a compliance period of length T. To prevent the permit price from exceeding the ceiling $\bar{p}$, the regulator then stands ready to sell at $\bar{p}$ additional permits from a cost containment reserve. Suppose the regulator asks a consulting economist to determine the smallest number of permits ($\bar{R}$) to hold in the reserve to defend the ceiling.

Such analyses implicitly assume that regulated firms surrender permits continually to cover their emissions. In that case, each firm abates to the point where any further reduction in emissions would cost more than merely purchasing a pollution permit instead to cover the marginal emissions. Suppose this implies that, at permit price p, emissions will be $E(p)$, where $E(\cdot)$ is a strictly decreasing function. Assume the economist knows this function.

creating a sawtooth pattern. In a discrete-time model where one period corresponds to one season, the price each period would be unchanging, and the rich intraseasonal dynamics would be concealed. Samuelson (1971) adopted this latter approach when discussing interseasonal carryovers under uncertainty.

Assume the economist also knows the interest rate r and recognizes that, for anyone to be willing to "bank" any of the $\bar{g}$ permits, given the absence of dividends, its price must grow at the rate of interest. Since the economist assumes that the regulated firms will surrender permits to cover contemporaneous emissions, he calculates that the price will initially be below the ceiling ($p_0 < \bar{p}$) and will reach it later—at $t_0 \in (0, T)$, where $p_0 e^{rt_0} = \bar{p}$ and $\bar{g} = \int_{x=0}^{t_0} E(p_0 e^{rx})dx$. He therefore reports that sales at the ceiling price from a reserve as small as $\bar{R}$ will keep the price from piercing the ceiling, where $\bar{R} = (T - t_0)E(\bar{p})$. That is, the economist predicts the price path $p(t) = \min(p_0 e^{rt}, \bar{p})$ for $t \in [0, T]$ as depicted in Figure 1, where the ceiling price is set at \$25 per ton CO_2, $\bar{p} = 25$.[9] Prior to t_0 (the 4^{th} quarter in the figure), regulated firms would use banked permits to cover their emissions; subsequent to t_0, firms would obtain the permits they need by purchasing at price $\bar{p}$ from the cost containment reserve.

Figure 1: An Analysis Assuming Continual Compliance Predicts that the Ceiling Will Not Be Pierced

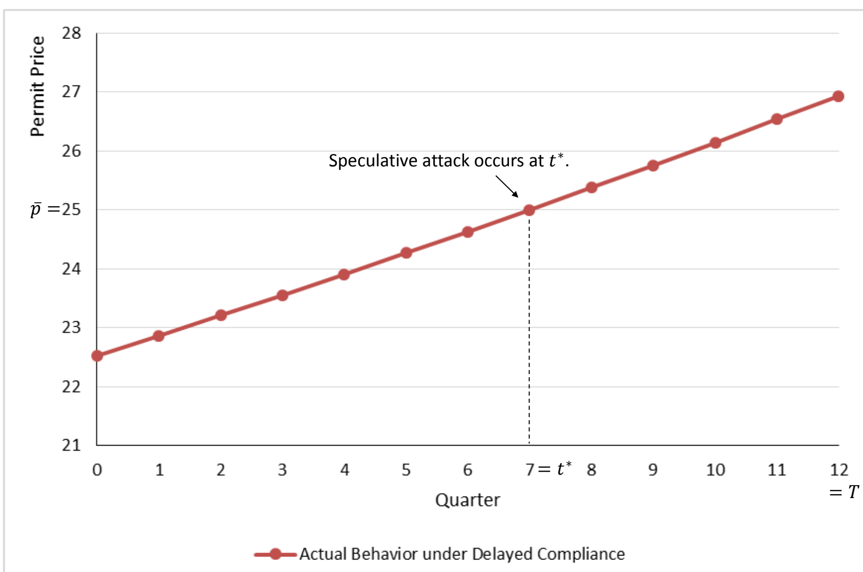


[9]We assume that a representative firm has a deterministic and static abatement cost function given by $c(e_t) = \frac{c_0}{2}(\bar{e} - e_t)^2$, where e_t is the firm's emissions at time t, $\bar{e}$ is the baseline emissions, and c_0 is the slope of the marginal cost function. Each time period corresponds to a quarter and the compliance period is three years (12 quarters), $t = 1, 2, \ldots, 12$. We choose the same parameter values that Fell and Morgenstern (2010) use, $c_0 = \$30$/ton per Gtons and $\bar{e} = 6.158$ (Gtons per year). Because each time period is one quarter long, we assume $\bar{e} = 6.158/4 = 1.5395$ for each period. We assume the quarterly interest rate is 1.5%.

However, since the actual program only requires firms to true-up at T, a reserve as small as $\bar{R}$ is too small to defend the ceiling. As we will see, price paths under delayed compliance rise throughout at the rate of interest. Hence, the price will initially fall below p_0 and will reach the ceiling $\bar{p}$ later than t_0.[10] When the price does reach the ceiling $\bar{p}$ at $t^* > t_0$, reserves in the cost containment reserve will appear ample since $\bar{R} > (T - t^*)E(\bar{p})$.

But contrary to the economist's prediction, the permit price will not pause, even momentarily, at the ceiling, but will swiftly surpass it even though the entire cost containment reserve is sold in a futile attempt to defend against this "speculative attack." Figure 2 illustrates the permit price path under delayed compliance; the attack occurs in the seventh quarter.[11]

Figure 2: The Predicted Price Path Under Delayed Compliance



Speculative attacks have been widely studied in other areas of economics but always assuming the analog of continual compliance.[12] A brief review of the mechanism underlying

[10]The actual equilibrium price path must cross the path forecasted by the analyst since cumulative emissions are the same ($\bar{g} + \bar{R}$) on each path; moreover, the crossing must occur on the horizontal part of the analyst's price path.

[11]The total number of permits issued or held in the cost containment reserve by the government is the same in both Figures 1 and 2 and, for the purpose of illustration, is determined so that the speculative attack occurs exactly at $t = 7$ under delayed compliance. The size of the reserve is set at the minimum level to defend the price ceiling under continual compliance in Figure 1.

[12]Salant and Henderson (1978) initiated the literature on speculative attacks under certainty and Salant

such precipitous equilibrium behavior under continual compliance may be useful and will explain why we extend the terminology to the delayed compliance context. Suppose some of the reserve $\bar{R}$ was instead shifted to the initial auction. Under continual compliance, an equilibrium with a speculative attack would inevitably result. For, suppose instead that the cost containment reserves were merely sold off at the steady rate $E(\bar{p})$. Then, given that $\bar{R}$ is the smallest reserve needed to sustain the ceiling, the reserves would be exhausted before T and bufferstock sales would necessarily jump down to zero. But this downward jump in sales would cause a contemporaneous upward jump in the permit price and this creates a foreseeable opportunity for unlimited profits. But this cannot occur in equilibrium. Otherwise every agent would seek to take advantage of it by buying on a massive scale immediately before the jump. In an equilibrium, the regulated firms or the "investment community" would purchase at an infinite rate the remainder of the regulator's stockpile in an instant—behavior dubbed a "speculative attack"—and then would draw down the acquired stock at a decreasing rate starting at $E(\bar{p})$.[13] As a result, there would be no upward jump in the price path.

If even more of the regulator's initial reserve was shifted to the initial auction, the ceiling would be defended for an even shorter interval before the speculative attack occurred under continual compliance. If the initial reserve is just small enough that the attack would occur as soon as the price ceiling is reached, the price path under continual compliance would *coincide* with the price path under delayed compliance.[14] That is why it seems reasonable to extend the term "speculative attack" to this precipitous behavior even though it occurs

(1983) extended it to uncertainty. Gaudet et al. (2002) showed that the same principles underlie naturally occurring "oil rushes" on common pools as well as "fishing derbies" when total allowable catch quotas are unallocated. Gaudet and Salant (2003) showed that the same principles also explain the sudden surge of H1A applications as immigration quotas fill as well as the sudden surge of imported durables as their quota fills. For discussions of how the same ideas were developed further in the international finance literature to explain currency crises, see Krugman (1999) and Flood et al. (2012).

[13] According to the Market Monitor Report (p. 6, 2014), although only 22% of allowances sold in the first 22 RGGI auctions went to speculators rather than to compliance entities and their affiliates, in the RGGI auction where the cost containment reserve was attacked, 55% of the allowances were purchased by speculators.

[14] The Holland and Moore (2013) sufficient condition would then hold.

under delayed compliance.

After the ceiling is pierced, the regulator will be unable to control prices during the remainder of the compliance period unless the regulator suddenly issues still more permits.[15] This is the situation in which RGGI now finds itself: "There are no more CCR allowances available for sale in 2014. It is therefore likely that the three auctions yet to come in 2014 will see even higher prices than this, since the CCR pressure relief valve will no longer be available" (Market Monitor Report, p. 1).

There is a significant chance that California's AB-32 will find itself in a similar situation in the foreseeable future.[16] Borentstein et al. (2014) calculate that there is a significant probability that the two lower tiers of the ceiling of AB-32 may be pierced during the next 8 years.[17] It is their political judgment that the regulator will find this situation intolerable: "It is highly unlikely that the political and regulatory process would allow the market to continue to operate freely at unduly high allowance prices, such as above the highest tier of the APCR" (Borenstein et al., p. 46).

One can imagine, therefore, the regulator asking some consulting economist (perhaps a different one) for the smallest number of *additional* permits ($\hat{g}$) to sell in a second auction at time t^* to keep the permit price from piercing the ceiling again during the *remainder* of the compliance period. The economist would recommend a second auction of $\hat{g} = \int_{x=0}^{T-t^*} E(\hat{p}e^{rx})dx - \bar{R}$ where $\hat{p}e^{r(T-t^*)} = \bar{p}$. Assuming the second auction is unanticipated, its price would fall below the ceiling and the permit price would grow at the rate of interest, reaching $\bar{p}$ at T. As a result, cumulative emissions would be $\hat{g}$ larger than if the regulator had conducted no second auction. Figure 3 illustrates the corresponding permit price

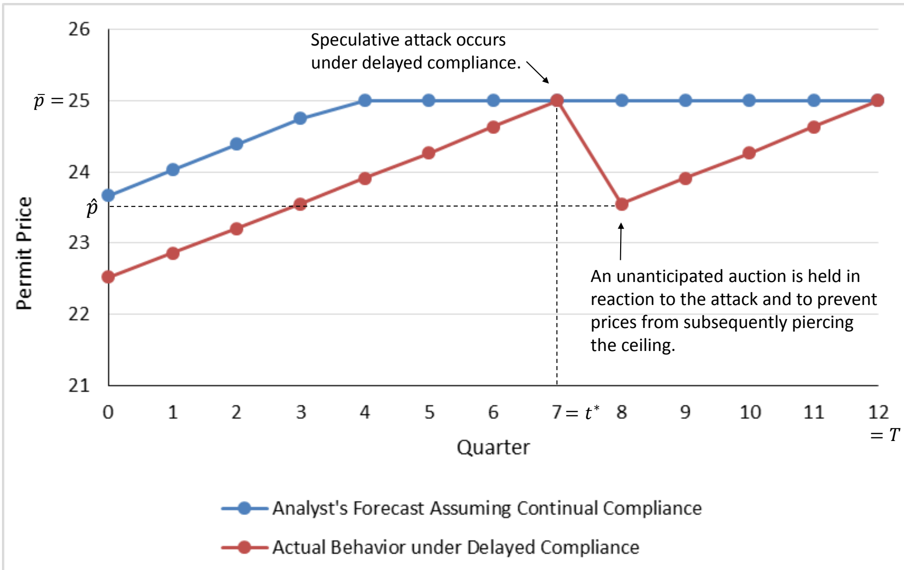
[15] The price path would be $p(t) = p'e^{rt}$ where p' uniquely solves $\int_{x=0}^{T} E(p'e^{rx})dx = \bar{g} + \bar{R}$.

[16] The RGGI cost containment reserve is implemented differently than AB-32's Allowance Price Containment Reserve. The former reserve supplements auctions if the settlement price exceeds the trigger; the latter sells reserves at fixed prices (although there are three rather than the one in our simplified exposition). As we will see in the next sections, however, if the same volume of reserves are available through either mechanism the cumulative supply curve will be the same and hence the equilibrium price path will be the same. The path of permit prices will rise at the rate of interest. Provided it crosses the trigger of the RGGI or the highest fixed price of AB-32, all of the reserves intended to keep prices down will be purchased by the private sector in a speculative attack.

[17] See their Figure 3.

path before and after the second auction and compares it with the path forecasted by the economic consultant assuming continual compliance.

Figure 3: An Unanticipated Second Auction To Protect the Ceiling from the Speculative Attack



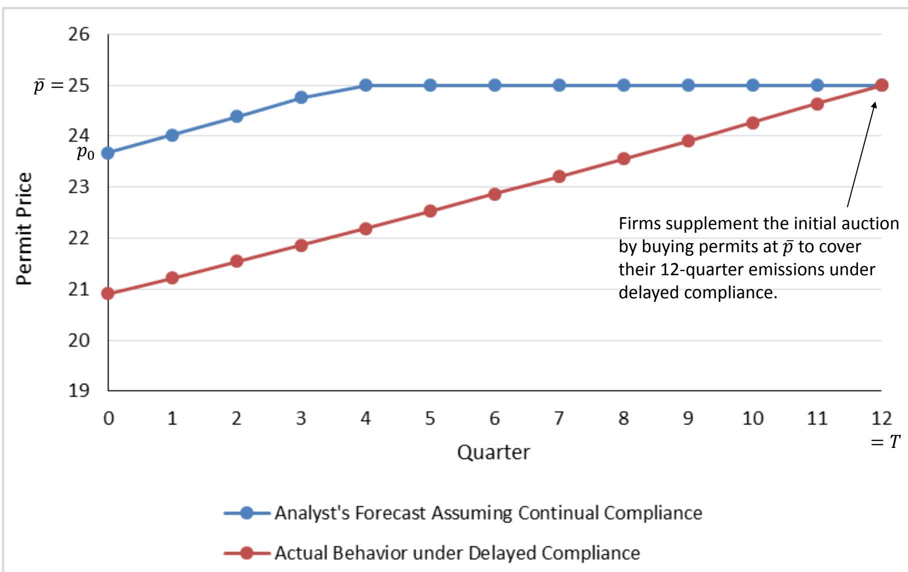
The economist erred in his original forecast because he implicitly assumed that differences in compliance timing do not matter. The price path he anticipated (Figure 1) in fact induces disequilibrium behavior under delayed compliance. Because it is cheapest (in present-value terms) to wait until the last instant to purchase permits to cover cumulative emissions, that is what every regulated firm would attempt to do; but there would then be insufficient permits in the reserve (or anywhere else) to satisfy this aggregate demand. If the economist realized that under delayed compliance permit prices must rise throughout at the rate of interest, he would have concluded that the smallest reserve that would prevent the ceiling from being pierced would result in a price path reaching the ceiling at T ($p(T) = \bar{p}$) as depicted in Figure 4. Such a path requires the much larger reserve $\tilde{R} = \int_{x=0}^{T} E(\bar{p}e^{-r(T-x)})dx - \bar{g} > \bar{R}$.[18]

If the regulator wants to prevent the price from piercing the ceiling, he can do so by stocking the cost containment reserve with sufficient reserves. To determine what size re-

[18]Cumulative emissions on the delayed compliance path would be larger since, as Figure 4 reflects, the permit price would be uniformly lower and there would be less incentive to abate.

serve is sufficient requires recognition that firms need not be in compliance between true-up dates. In our example, determining the proper size reserve is relatively easy. But the exercise becomes more difficult if interim auctions with reserve prices are also involved since one must determine how many permits will be injected from these sources as well. It is, therefore, important to develop a general methodology to analyze the equilibrium under delayed compliance. In the following sections, we describe that methodology and illustrate its application.

Figure 4: The Price Path with the Smallest Reserves Sufficient to Defend the Ceiling Under Delayed Compliance



3 A Methodology for Analyzing Cap-and-Trade Programs under Delayed Compliance

Under delayed compliance, prices can never rise faster than the rate of interest; otherwise traders would attempt to buy low and sell high on an infinite scale. Nor can prices ever rise *slower* than the rate of interest. For if they did, the highest capitalized price would *strictly exceed* the lowest capitalized price. But then everyone with an initial allocation of permits would want to sell them at the highest capitalized price and there would be no one on the other side of the market willing to buy permits at that price; as a result, there would be massive excess supply.[19] It follows that in any equilibrium under delayed compliance, prices must rise throughout the compliance period at the rate of interest. Anticipated government auctions or sales from a finite reserve at a fixed price will not slow this rate of price appreciation although, as we will show, they will determine the position of the price path or, equivalently, the permit price prevailing at the compliance date.

We assume throughout that no borrowing is allowed within a compliance period. If there are future compliance periods, we assume that permits cannot be borrowed from them and that the demand and supply conditions prevailing in each of them is the same as in the current compliance period; hence, the price path in the current compliance period will repeat in each future compliance period and there will be no incentive to carry permits into the next compliance period.[20] The equilibrium permit price at the compliance date equates the demand for permits required to cover the cumulative emissions which have occurred since the last compliance date with the cumulative supply of permits consisting of the initial allocations *plus* the subsequent injections of additional permits during the compliance period. Since the equilibrium price path under delayed compliance must rise at the rate of interest, we may index such paths by the price expected to prevail at the date of compliance. Denote that price as $\mathbf{P}$.

To determine the permit price at the compliance date, we proceed as follows. For each possible terminal price, determine the (unique) associated price path over the compliance period. To determine the cumulative demand for permits along that path, note that at every instant firms will abate up to the point where their marginal cost of abatement capitalized

[19] This argument implicitly assumes that private agents are the only purchasers of permits. We assume that the government never purchases permits since none of the delayed compliance programs we consider envision that. If the government did purchase permits, the equilibrium price path under delayed compliance could rise slower than the rate of interest.

[20] We assume that banking is allowed without limitation. Our stationarity assumptions imply that, although banking would occur within each compliance period, carrying permits from one compliance period to the next would be strictly unprofitable. The algorithm which follows in the text can be easily modified, however, if nonstationarities made banking of permits between compliance periods profitable.

to the compliance date equals the permit price anticipated to prevail at that date. Compute the aggregate cumulative emissions of the regulated entities over time. Firms will need a matching number of permits at the compliance date. This procedure provides one price-quantity pair on the cumulative demand curve for permits. Repeat the procedure to generate the other points on the demand curve.[21]

Deriving the cumulative supply of permits as a function of the terminal price is somewhat trickier. We assume that firm i ($i \in \{1, \ldots, n\}$) can reduce its emissions per unit time to $e_i(t)$ at time t by abating at cost $c_i(e_i(t))$, where firm i's cost is a strictly decreasing, strictly convex, differentiable function of emissions. To avoid corners, we assume the Inada condition holds: $-c_i'(e) \to \infty$ as $e \to 0$. Moreover, at a sufficiently high level of emissions ("baseline emissions," $\bar{e}_i$), the firm's cost declines to zero and approaches that level at a zero slope: $c_i(\bar{e}_i) = c_i'(\bar{e}_i) = 0$. Then firm i chooses its emissions path $e_i(t)$ to minimize its total cost of complying with the cap-and-trade regulation. It minimizes $c_i(e_i(t))e^{r(T-t)} + e_i(t)\mathbf{P}$, where T is the compliance date.[22] Its optimal emissions path therefore solves:

$$-c_i'(e_i(t)) = \mathbf{P}e^{r(t-T)}, \text{ for } t \in [0, T] \text{ and } i = 1, \ldots n. \tag{1}$$

Given the properties of the n cost functions, the emissions of each firm at any instant are a continuous, strictly decreasing function of $\mathbf{P}$. From the emissions paths of the firms, we

[21] To simplify the exposition, we assume that firms do not abate by investing in new or altered technology. Otherwise a firm's current abatement decision would affect its future cost of emissions, and each firm would have a dynamic investment problem to solve. Accounting for this would complicate the derivation of cumulative emissions if the permit price rises at the rate of interest and hence the aggregate cumulative demand for permits, but it would not alter any of our points. The equilibrium price path under delayed compliance would still be determined at a terminal price that equates cumulative supply to the altered cumulative demand, and this path would fail to equilibrate markets under continual compliance for the reasons we discuss. We have chosen, therefore, to abstract from this real world complication. As noted in the concluding section, one cannot abstract from this complication when conducting welfare analysis.

[22] Note that, under continual compliance, firm i chooses its emissions $e_i(t)$ at time t so as to minimize the sum of the abatement cost and the cost of purchasing permits at time t, $c_i(e_i(t)) + e_i(t)p(t)$. The firm i's optimal emission path under continual compliance is a function of the current price $p(t)$ but not $\mathbf{P}$ and solves $-c_i(e_i(t)) = p(t)$. Therefore, the aggregate contemporaneous emission function under continual compliance, which is denoted as $E(p)$ in the previous section, can be expressed as $E(p(t)) = \sum_{i=1}^{n} e_i(t)$ such that $-c_i(e_i(t)) = p(t)$ while the emission path for firm i under delayed compliance is determined by equation (1).

can determine the cumulative aggregate demand for permits through time τ as a function of $\mathbf{P}$:

$$D(\mathbf{P},\tau) = \int_{t=0}^{\tau} \sum_{i=1}^{n} e_i(t)dt \text{ for } \tau \in [0,T]. \tag{2}$$

$D(\mathbf{P},\tau)$ is continuous, strictly decreasing in its first argument and strictly increasing and strictly concave in its second argument. The intercepts are $D(\mathbf{P},0) = 0$ and $D(0,\tau) = \tau \sum_{i=1}^{n} \bar{e}_i$. We will make extensive use of this function in the subsequent analysis.

For any particular government method of injecting permits, we can define $S(\mathbf{P},\tau)$ as the government's cumulative supply of permits until time τ on a price path rising at the rate of interest and ending at $\mathbf{P}$. Under delayed compliance, any price path such that $D(\mathbf{P},T) = S(\mathbf{P},T)$ equilibrates the market.

Holland and Moore (2013) prove that the paths of permit prices and emissions are invariant to compliance timing under a strong sufficient condition. Under that assumption, analysts assuming continual compliance will nonetheless correctly predict the paths of prices and emissions even though, as Table 1 reflects, all trading schemes in fact utilize delayed compliance.

We provide instead a condition to determine when predictions using continual compliance will inevitably introduce error. To do so, we ask under what circumstances will the equilibrium under delayed compliance generate a disequilibrium under continual compliance.

Continual compliance requires not merely that that $D(\mathbf{P},T) = S(\mathbf{P},T)$ but also that $D(\mathbf{P},\tau) \leq S(\mathbf{P},\tau)$ for *all* $\tau \in [0,T)$. That is, in the continual compliance regime, agents must be provided enough permits to be able to cover their emissions at *every* instant and not merely the last one. Since the requirement of equilibrium is more restrictive under continual compliance, price paths that equilibrate the market under delayed compliance but where $D(\mathbf{P},\tau) > S(\mathbf{P},\tau)$ for some $\tau \in [0,T)$, fail to equilibrate it under continual compliance.

4 Applications of the Methodology

We now illustrate how to determine the paths of prices and emissions when some permits are allocated initially and others are injected during the compliance period by auctions with or without reserve prices. As Table 1 confirms, this method of injecting permits has been used as far back as the Acid Rain Program.

4.1 Auctions with Reserve Prices

Assume that g permits are "grandfathered" at the outset and that the number grandfathered is smaller than the cumulative emissions that would have occurred without a cap-and-trade program ($g < T\sum_{i=1}^{n}\bar{e}_i$).

Assume that the government commits at the outset to conduct a sequence of auctions. The date, amount, and reserve price of each auction is announced at the outset. Let t_i denote the date of the i^{th} auction, a_i its amount, and $\underline{p}_i$ its reserve price (assumed strictly positive) for $i = 1, \ldots A$, where A is the total number of auctions to be held during the compliance period, $[0, T]$.

To determine the equilibrium price path under delayed compliance, we construct the cumulative demand and cumulative supply curves and determine their unique point of intersection. The cumulative demand curve is simply $D(\mathbf{P}, T)$, which is downward-sloping with respect to $\mathbf{P}$. The cumulative supply curve $S(\mathbf{P}, T)$ is a step-function. For the price path with the terminal price of zero, aggregate supply consists of the g grandfathered permits. As the terminal price is increased, it eventually equals lowest capitalized reserve price. At that terminal price, the cumulative supply is indeterminate—as small as g and as large as g plus the amount offered at the auction with the lowest capitalized reserve price. If the terminal price is slightly higher, the cumulative supply equals the upper end of this interval. Cumulative supply would remain at that level until the terminal price reached the next-to-the-lowest capitalized reserve price. A sufficiently high terminal price will equal the highest of the capitalized reserve prices of the A auctions. Any higher terminal price will elicit the maximal supply of $g + \sum_{i=1}^{A} a_i$ permits.

There exists a unique equilibrium price path and terminal price, $\mathbf{P}$. Existence follows since a zero terminal price would generate excess cumulative demand (by assumption, $T \sum_{i=1}^{n} \bar{e}_i > g$) while a sufficiently high terminal price would generate excess cumulative supply (cumulative supply $g + \sum_{i=1}^{A} a_i$ is bounded away from zero and cumulative demand approaches zero for sufficiently high $\mathbf{P}$). Moreover the intersection point must be unique since, at any higher price, demand is strictly smaller and supply weakly larger while, at any lower price, demand is strictly larger and supply weakly smaller.

To construct the supply curve geometrically, proceed as follows: (1) on a diagram with time on the horizontal axis and price per permit on the vertical axis (see Figure 5), record the date and reserve-price pair $(t_i, \underline{p}_i)$ of each of the A auctions; (2) determine the capitalized value ($\mathbf{P_i}$) of each reserve price by drawing through each of these A points a price path rising at the rate of interest and noting its height at T ($\mathbf{P_i} = \underline{p}_i e^{r(T-t_i)}$). For terminal prices smaller than the smallest capitalized reserve price, only the g grandfathered permits are supplied to the market. For higher prices, the cumulative supply function $S(\mathbf{P}, T)$ will have a horizontal step of length a_i at height $\mathbf{P_i}$ for $i = 1, \ldots A$.

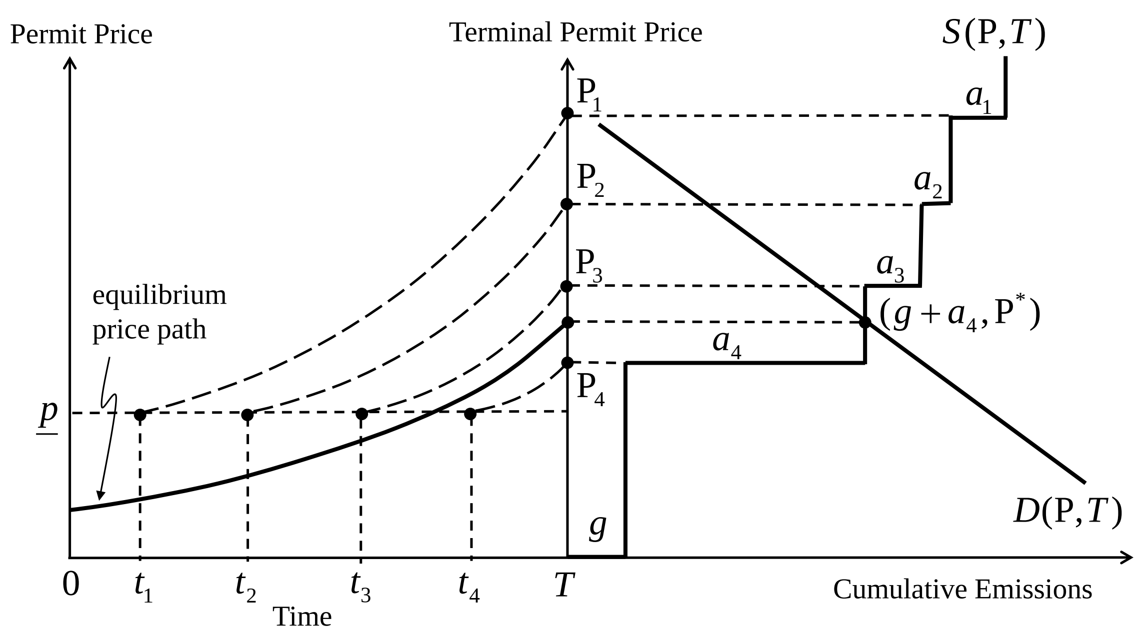
Our methodology can be used to predict the consequences of *any* exogenous path of auction reserve prices. For example, suppose the auction reserve price rises exactly at the rate of interest. Then, if bids at one auction strictly exceed its reserve price, bids at the other auctions will strictly exceed their reserve prices. Conversely, if no bids meet the reserve price in one auction, none will meet it in any other auction.

If instead the exogenous auction reserve price rises faster than the rate of interest, then if bids fail to meet the reserve price in one auction, no bids will be acceptable in subsequent auctions while if bids *do* meet the reserve price in one auction, bids in prior auctions will also be acceptable. Consequently, when the exogenous reserve price rises faster than the rate of interest, auctions where bids fail to meet the reserve price cluster at the end of the compliance period.

If instead the exogenous auction reserve prices rise slower than the rate of interest, then if bids fail to meet the reserve price in one auction, no bids will be acceptable in earlier auctions while if bids do meet the reserve price in one auction, bids in subsequent auctions will also be acceptable. Consequently, when the exogenous reserve price rises slower than the rate of interest, auctions where bids fail to meet the reserve price cluster at the beginning of the compliance period.

Although our methodology can be applied to any exogenous path of reserve prices, we illustrate it below in the simplest manner. Hence, we assume that all the reserve prices are the same. In Figure 5, all auctions have the same reserve price ($\underline{p}_i = \underline{p}_j$). Since these reserve prices rise by less than the rate of interest, $\mathbf{P_1} > \mathbf{P_2} > \mathbf{P_3} > \mathbf{P_4}$. In the example portrayed, the equilibrium terminal price $\mathbf{P}^*$ is contained in the open interval $(\mathbf{P_4}, \mathbf{P_3})$. Hence, no bids are accepted at the first three auctions but all of a_4 permits are sold at the fourth auction at the price $\mathbf{P}^* e^{r(t_4 - T)} > \underline{p}$. Therefore, in equilibrium emissions equal $g + a_4$.

Figure 5: The Cumulative Demand and Supply, and the Equilibrium Price Path in the Case of Reserve Price Auctions

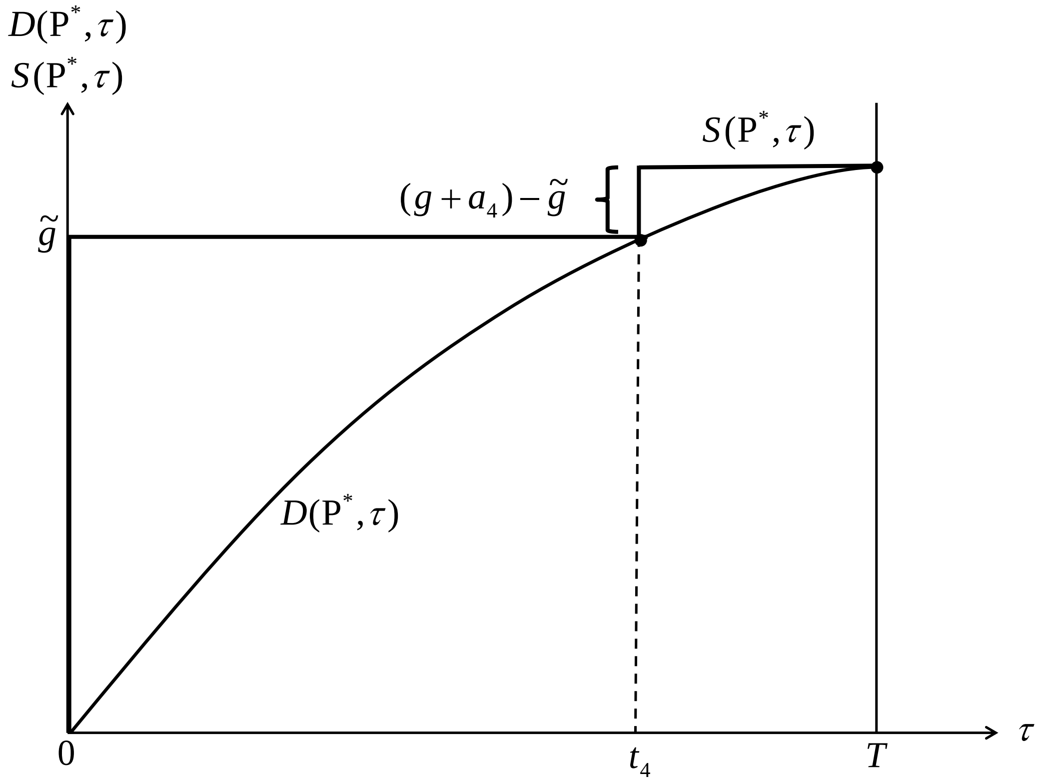


If the government had auctioned no permits at t_4 but had instead added these a_4 permits to the number grandfathered, then the cumulative supply curve would become $g + a_4$ for terminal prices below $\mathbf{P_4}$ but the modified cumulative supply curve would still intersect the unchanged cumulative demand curve at the same point. Hence, the equilibrium price path would not change under delayed compliance nor would the cumulative emissions it induces. Alternatively, if the government had grandfathered no permits but had instead auctioned these g permits along with the a_4 permits at t_4, then the cumulative supply at prices below $\mathbf{P_4}$ would be zero and the cumulative supply at $\mathbf{P_4}$ would be as large as $g + a_4$. Nonetheless this modified cumulative supply curve would still intersect the cumulative demand curve at the same point. Neither change would affect the equilibrium under delayed compliance.

If the entire sum of permits $(g+a_4)$ was grandfathered at the outset, then this same price path would also equilibrate the market even if firms were required to cover their emissions continually during the compliance period. But if all of these permits were made available instead at t_4, then an analyst erroneously assuming continual compliance would introduce errors into his forecast since, for some $\tau < T, D(\mathbf{P}, \tau) > S(\mathbf{P}, \tau)$. In Figure 6, we plot cumulative supply and demand until τ along the equilibrium price path under delayed compliance (the price path ending at $\mathbf{P}^*$). Since the path generates an equilibrium under delayed compliance, the two curves intersect at T and $D(\mathbf{P}^*, T) = g + a_4$. An equilibrium under continual compliance would require in addition that the cumulative supply curve lies nowhere strictly below the cumulative demand curve for $\tau \in [0, T)$. Let $\tilde{g}$ be the number of permits the government chooses to grandfather initially. We have drawn the boundary case where $\tilde{g} = D(\mathbf{P}^*, t_4)$ permits are grandfathered and $(g + a_4) - \tilde{g}$ are auctioned at t_4. If the government grandfathered strictly less than $D(\mathbf{P}^*, t_4)$ permits and added them instead to the amount auctioned at t_4, then an analyst assuming that firms were continually in compliance between true-up dates would introduce errors into his forecast.[23]

[23]In particular, he would erroneously predict that the price path would consist of segments rising at the rate of interest, separated by a downward jump at t_4.

Figure 6: The Boundary Case in the Reserve Price Auctions



4.2 Non-existence of Equilibrium Induced by the Rules of California's AB-32

As we have seen, our methodology can be used to predict the paths of emissions and prices under delayed compliance and to determine how much will be sold in each auction and at what price. It can also be used to identify problematic features of these sometimes arcane regulations. To illustrate, we consider a strange provision in California's AB-32 under which *some* of the permits which did not sell in earlier auctions may be offered in later auctions *provided* specified conditions are met.

Suppose that cumulative demand was so large that $\mathbf{P}^* \in (\mathbf{P_2}, \mathbf{P_1})$. That is, every auction after t_1 sells out, but bids in this first auction are below its reserve price ($\underline{\mathrm{p}}_1$). Under the rules of California's AB-32, the a_1 permits which failed to sell in the first auction would be returned to the "Auction Holding Account."[24] Some of these permits would be available for sale in the fourth auction since it would have occurred "after two consecutive auctions have resulted in an auction settlement price greater than the applicable Auction Reserve Price."[25] However, not all of the a_1 permits could be made available. At most, the number of permits which can be added to the auction at t_4 is 25% of a_4.[26] At the price that would equate cumulative demand and supply in the absence of these additional permits, excess supply would occur because these unsold additional permits would be offered in an auction where the settlement price would strictly exceed the reserve price. As a result, if any equilibrium exists, it will occur at a lower terminal price.

Offering unsold permits for sale if and only if permits are sold at two preceding auctions in a row can create a situation, however, where no competitive equilibrium exists. Suppose, for example, that every bid is strictly below the reserve price in the first auction but the settlement prices in the next two auctions strictly exceed their respective reserve prices. Suppose cumulative demand is sufficiently high that in the absence of the rule regarding unsold permits that $\mathbf{P}^* \in (\mathbf{P_2}, \mathbf{P_1})$. Under this rule $\min(a_1, 0.25a_4)$ of the permits from the first auction can be offered in the fourth auction. If $\min(a_1, 0.25a_4) \geq D(\mathbf{P_2}) - g - a_2 - a_3 - a_4$, then there will be excess supply at any terminal price strictly exceeding $\mathbf{P_2}$. But at any terminal price equal to or strictly below $\mathbf{P_2}$, there will be excess demand since, in the absence of two consecutive auctions where permits are sold at prices strictly higher than the reserve price, *none* of the unsold permits from auction 1 can be offered for sale in the fourth auction, and then $D(\mathbf{P}) > g + a_2 + a_3 + a_4$ holds for all $\mathbf{P} \leq \mathbf{P_2}$. We illustrate a situation with no equilibrium in Figure 7.

To summarize in words, if no permits from auction 1 are sold in auction 4, then the equilibrium price path would have to terminate strictly above $\mathbf{P_2}$. But if it did, then this provision of AB-32 would *mandate* the sale of permits from auction 1, contrary to the

[24] Final Regulation Order, §95911. Format for Auction of California GHG Allowances. (b) (4) (A)

[25] Quoted from Final Regulation Order, §95911. Format for Auction of California GHG Allowances. (b) (4) (B)

[26] Final Regulation Order, §95911. Format for Auction of California GHG Allowances. (b) (4) (C)

premise. On the other hand, if permits from auction 1 *are* offered in auction 4, then they would all sell since, in this example, the market price of permits would far exceed auction 4's reserve price. But then one can easily construct cases where the additional permits from auction 1 would result in a price path terminating at or strictly below $\mathbf{P_2}$. In that case, there would not be two consecutive auctions with auction settlement prices strictly above their respective reserve prices, and the same provision would *prohibit* sale of permits from auction 1, again contrary to our premise. Hence, this provision of AB-32 would again undermine the establishment of an equilibrium. Since there are only two possible cases to consider and both lead to disequilibrium, no equilibrium exists in our example under this provision of AB-32.

To insure that an equilibrium *does* exist, the regulation needs to be revised. One solution would be to *prohibit* offering permits at auction more than once. If, however, permits from auctions where the settlement price was strictly below the reserve price *are* allowed to be sold later, they should be offered later *regardless* of what happens in other auctions.

Suppose, for example, that unsold permits from earlier auctions could be offered at the last auction in the compliance period as long as their sum did not increase the amount offered in that auction by more than 25%. This has much the same spirit as the provision it would replace. But under this proposed new rule, no disequilibrium would result. As we will see, the equilibrium will have one of the following characteristics: the settlement price of auction 1 alone is strictly below its reserve price; the settlement prices of auctions 1 and 2 are strictly below their respective reserve prices; the settlement prices of auctions 1, 2, and 3 are strictly below their respective reserve prices; or so many permits are grandfathered initially that in equilibrium nothing is sold in any auction, even in auction 4. To see which situation occurs, we consider each case in turn.

If no permits from auction 1 sold, then, under the proposed new rule, they must be offered again in auction 4. Hence, the step in the cumulative supply "function" (correspondence, really) at height $\mathbf{P_4}$ would lengthen to $a_4 + \min(a_1, .25a_4)$. And the horizontal step at height $\mathbf{P_1}$ would be removed. Cumulative supply and demand would intersect. If the intersection occurred at $\mathbf{P_2}$ or above, then we would have found the terminal price in the equilibrium and hence the entire price path.

If the intersection occurred strictly below $\mathbf{P_2}$, however, then none of the permits in auction 2 would have sold either, and the new rule requires that they be made available in auction 4 as well. In that case, the step at height P_4 in the cumulative supply function would lengthen to $a_4 + \min(a_1 + a_2, .25a_4)$ and the steps at height $\mathbf{P_1}$ and $\mathbf{P_2}$ would be removed. Cumulative supply and demand would again intersect. If the intersection occurred at $\mathbf{P_3}$ or above, then we would have found the terminal price in the equilibrium and hence the entire price path.

If the intersection occurred strictly below $\mathbf{P_3}$, however, then none of the permits in auction 3 would have sold either, and the new rule requires that they be made available in auction 4 as well. In that case, the step in the cumulative supply function at height $\mathbf{P_4}$ would lengthen to $a_4 + \min(a_1 + a_2 + a_3, .25a_4)$ and the steps at height $\mathbf{P_1}, \mathbf{P_2}$ and $\mathbf{P_3}$ would be removed. Cumulative supply and demand would again intersect. If the intersection occurred at $\mathbf{P_4}$ or above, then we would have found the terminal price in the equilibrium and hence the entire price path.

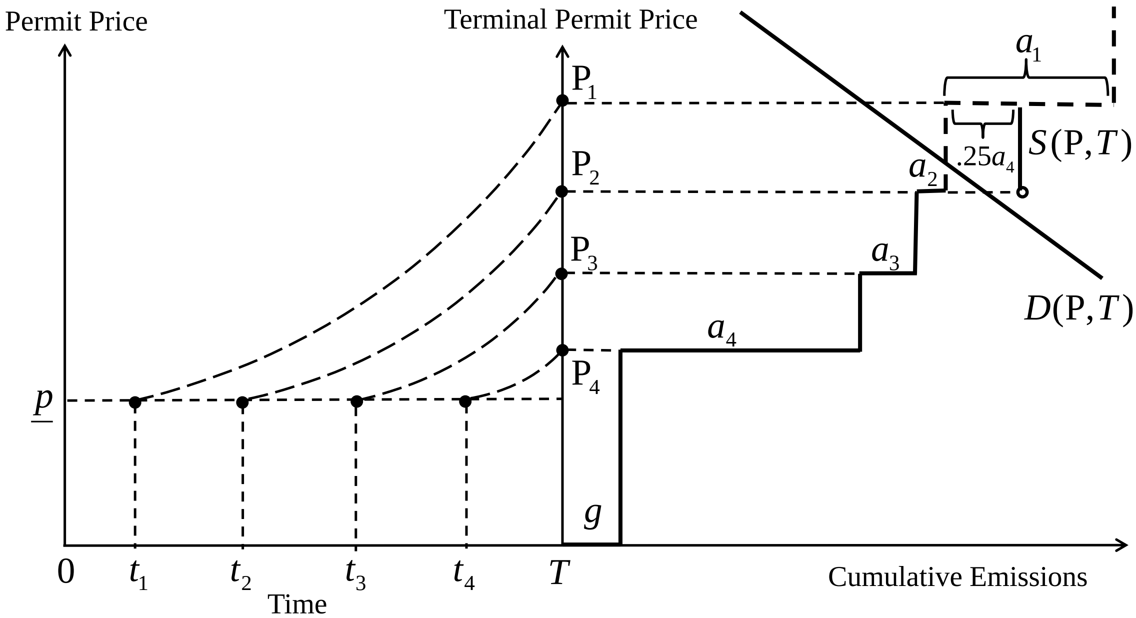
The only other possibility is that so many permits were grandfathered initially that in the equilibrium nothing sells at any auction. In that case, the terminal price will be where the cumulative demand curve intersects the vertical segment of the cumulative supply curve at horizontal distance g.

Thus, the proposed new rule would allow a limited number of permits from auctions where nothing sold to be offered in the final auction of the compliance period. But unlike the current provision in AB-32, the proposed rule would never undermine equilibrium.

4.3 Sales at Specified Prices

Permits can also be injected during the compliance period by sales at a specified price, which we denote $\bar{p}$. In section 2, we calculated the smallest reserve ($\bar{R}$) necessary to prevent the

Figure 7: The Case of No Competitive Equilibrium under the Rules of California's AB-32

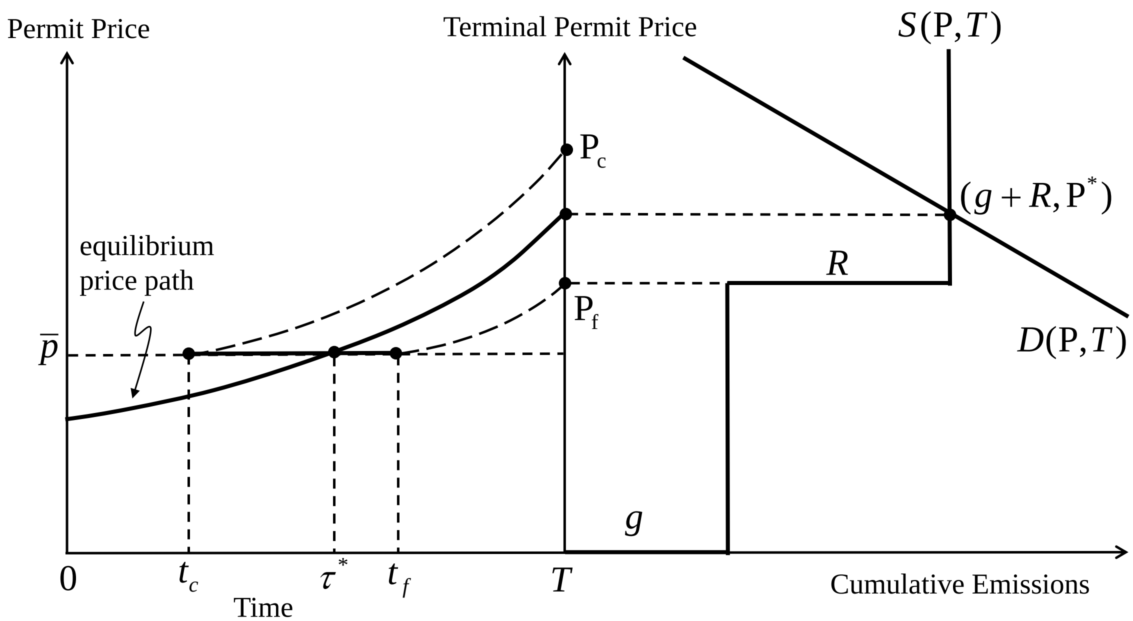


price from exceeding the ceiling $\bar{p}$. It is straightforward to adapt that analysis if there is also a sequence of interim auctions, each scheduled at a different date and with its own size and reserve price. Under delayed compliance, the equilibrium price path will rise at the rate of interest, reaching $\mathbf{P} = \bar{p}$ at T. Compute $D(\bar{p}, T)$ along this path. Compute $S(\bar{p}, T)$ under the assumption that the cost containment reserve is empty. If a cost containment reserve is necessary to keep the price from piercing the ceiling, then $D(\bar{p}, T) - S(\bar{p}, T) > 0$ and this difference is the smallest amount that must be stocked in the cost containment reserve to keep the price from piercing the ceiling.

Henceforth, we assume in this section that all permits not grandfathered at the outset are injected by such sales. Such sales can occur over a specified time interval which commences at t_c and finishes at t_f or until all of the R permits in the "Cost Containment Reserve" have been sold. The Kerry-Lieberman bill envisioned such sales over a finite interval. They resemble a continuum of auctions with reserve price $\underline{p}$ over the time interval $[t_c, t_f]$ but with R available in the initial auction, and *everything* unsold in one auction immediately available for sale in subsequent auctions.

Since the sales price over the interval is constant, the price at t_f has the smallest capitalized value ($\mathbf{P_f} = \bar{p}e^{r(T-t_f)}$). In Figure 8, we depict the interval of offers and the sales price. As in Figure 5, we depict $\mathbf{P_f}$ by noting the terminal price on the path through the point $(t_f, \bar{p})$ rising at the rate of interest. To derive the cumulative supply curve, note that if the terminal price is strictly smaller than $\mathbf{P_f}$, then nothing would sell during this time interval and the cumulative supply would just be g. If the terminal price is strictly larger, then the cumulative supply would be $g + R$. If the terminal price is exactly $\mathbf{P_f}$ then the cumulative supply is any number of permits in the closed interval $[g, g + R]$.

Figure 8: The Cumulative Demand and Supply, and the Equilibrium Price Path in the Case of Sales at Fixed Prices



Suppose cumulative demand is sufficiently large that under delayed compliance the terminal price strictly exceeds $\mathbf{P_f}$. Then R permits sell instantaneously, either at some interior date $\tau^* \in (t_c, t_f)$ or at the first moment of the sale (t_c), where τ^* is determined so that $\mathbf{P}^* e^{-r(T-\tau^*)} = \bar{p}$ for $\mathbf{P_f} < \mathbf{P}^* < \mathbf{P_c}$. In either case, a "speculative attack" would occur in which purchases occur at an infinite rate.[27] The speculative attack occurs as the ceiling price is crossed. The permits are then held by private agents until the compliance date, when they

[27]For a discussion of speculative attacks, see footnote 12 and the associated text.

are returned to the government to cover emissions since the end of the previous compliance period.

Suppose in the equilibrium that the terminal price is $\mathbf{P}^*$ and cumulative emissions are $g + R$. Hence, $D(\mathbf{P}^*, T) = g + R$. Suppose the speculative attack occurs at the interior point $\tau^* > t_c$. Reallocating the $g + R$ permits between the initial allocation and the Cost Containment Reserve will not alter the equilibrium price path or the date of the attack.

When a speculative attack occurs ($\mathbf{P}^* > \mathbf{P_f}$), marginally changing the level of the price ceiling $\bar{p}$ does not affect the equilibrium price path and only alters the date of the attack under delayed compliance.

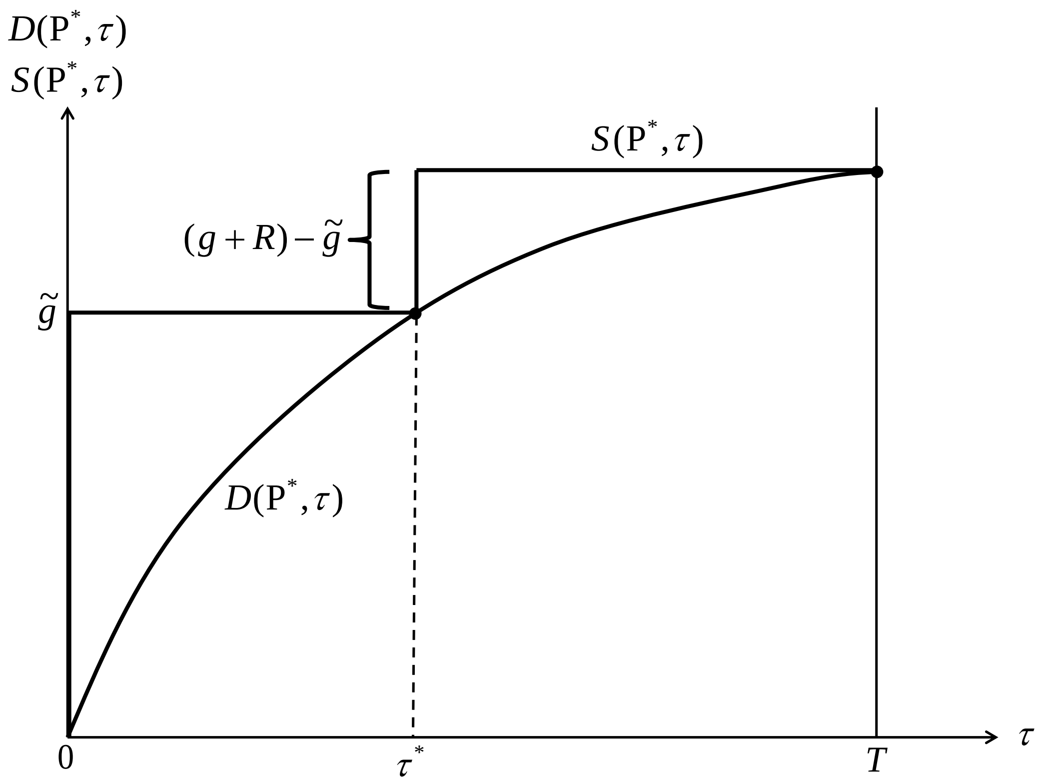
In Figure 9, we depict the boundary case where the government chooses to grandfather $\tilde{g} = D(\mathbf{P}^*, \tau^*)$ and to stock the cost containment reserve with the remaining permits. Hence, $g + R - \tilde{g} = D(\mathbf{P}^*, T) - D(\mathbf{P}^*, \tau^*)$. If $\tilde{g}$ were reduced so that more permits were moved from the initial allocation to the Cost Containment Reserve, then an analyst assuming that firms must be continually in compliance between true-up dates would introduce errors into his forecast.[28]

5 Conclusion

Our goal in this paper has been to improve dynamic analyses of cap-and-trade programs by identifying an erroneous assumption implicit in them and showing how it can be replaced by an appropriate one. Contrary to what has been assumed in previous analyses, cap-and-trade programs never require firms to be in compliance between true-up dates. Analyses making this erroneous assumption are prone to error when, as most programs do, they also mandate interim injections through auctions or safety-valve sales. We propose a straightforward methodology for analyzing such programs which dispenses with the erroneous assumption about compliance timing. We then show how it can be applied even when the complex

[28]In particular, he would predict that the price path would have a segment rising at the rate of interest until t_c and then (weakly) dropping to $\bar{p}$ for an endogenous interval of time before rising continuously from $\bar{p}$ again at the rate of interest.

Figure 9: The Boundary Case in the Sales at Fixed Prices



features of real-world regulations are taken into account.

To make our points clearly, we have assumed away all forms of uncertainty. This has permitted us to explain issues analytically that might have been obscured if we were confined to dynamic stochastic simulations. In the future, however, we plan to take account of uncertainty using such simulations.

Permit markets may be subject to three kinds of uncertainty: (1) uncertainty about the aggregate demand for permits that will be resolved by an information disclosure at a fixed date in the future; (2) aggregate demand shocks in each period; and (3) regulatory uncertainty.

The consequences of disclosing information at a known time about the demand for permits is illustrated by the collapse of the permit price in Europe following the disclosure of low demand for permits. In the case of demand shocks each period, the price path would become stochastic rather than deterministic. But if agents are risk neutral, little would change. If one works backward from the compliance period, assuming that on that date (1) all permits will be surrendered to the government and (2) agents with uncovered cumulative emissions must pay a well-specified penalty then, to equilibrate the market in the penultimate period under delayed compliance, the penultimate price must equal the discounted price expected in the next period. For, if that expected price were strictly higher, there would be excess demand for permits in the current period; and, if that expected price were strictly lower, there would be excess supply in the current period. But the same argument can be repeated as one works backward. In the stochastic equilibrium under delayed compliance, therefore, the price in every period must *equal* the discounted price then expected to prevail in any future period.

Regulatory uncertainty arises in part from the government's understandable goal of having the flexibility to cope with future circumstances. Regulators tend to avoid committing to future actions or policy rules. They prefer "discretion" to "precommitment." However, government flexibility, while understandable, distorts the intertemporal decision-making of private agents. This is true whether the private agents fully understand the regulator's objectives (Kydland and Prescott, 1977) or regard government actions as somewhat random (Salant and Henderson, 1978). McWilliams et al. (2011) have shown the importance of regulatory uncertainty in one permit market. Participants in the SO_2 market anticipated that at some unknown time in the future more permits would be required to cover each unit of SO_2 emissions and the price of permits would jump up. Anticipation of this uncertain event resulted in higher permit prices and more abatement; moreover, agents were willing to hold permits even though the permit price was rising by less than the rate of interest because of the capital gain they would receive when the uncertainty was resolved.

An important topic left for future work is welfare analysis. Under our assumptions, the cumulative emissions that arise in the equilibrium will be generated at least discounted costs since the marginal cost of abatement has the same present value in every period. However, under the plausible assumption that the stock of greenhouse gasses generates a flow of damages at each point in time, it has been shown (Kling and Rubin, 1997; Leiby and Rubin, 2001) that such an emissions path does not minimize the more relevant welfare functional—*the discounted sum of damages plus abatement costs.* In determining the socially optimal emissions path given such a damage function and in quantifying the welfare loss that would occur under a delayed compliance regime, we will have to take explicit account of when firms install their abatement technologies.[29]

Once we have calculated the welfare consequences of periods of delayed compliance of any given length, we can formally assess the optimal length of each compliance period. But some observations need not await formal analysis. Intuitively, the longer is each compliance period, the less likely is the government to wait for a period to end before intervening. In addition, the longer the compliance period, the greater the chance that (1) firms with uncovered pollution will go out of business before having to comply and (2) large utilities which have not complied before the end of the compliance period will evade regulation by threatening to shut down if forced to comply.

References

[1] Borenstein, Severin, James Bushnell, Frank A. Wolak, and Matthew Zaragoza-Watkins. 2014. "Expecting the Unexpected: Emissions Uncertainty and Environmental Market Design." Working Paper.

[2] Burtraw, Dallas, Karen Palmer, and Danny Kahn. 2010. "A Symmetric Safety Valve." *Energy Policy* 38(9), pp. 4921-4932.

[3] California Air Resources Board. 2011. "Final Regulation Order (corrected on December 21, 2011), Article 5: California Cap on Greenhouse

[29]See footnote 21.

Gas Emissions and Market-Based Compliance Mechanisms.” available at http://www.arb.ca.gov/regact/2010/capandtrade10/finalrevfro.pdf.

[4] Cronshaw, Mark B. and Jamie B. Kruse. 1996. “Regulated Firms in Pollution Permit Markets with Banking.” *Journal of Regulatory Economics* 9(2), pp. 179-189.

[5] Fell, Harrison, Dallas Burtraw, Richard D. Morgenstern, and Karen L. Palmer. 2012a. “Soft and Hard Price Collars in a Cap-and-Trade System: A Comparative Analysis.” *Journal of Environmental Economics and Management* 64(2), pp. 183-198.

[6] Fell, Harrison, Ian A. MacKenzie, and William A. Pizer. 2012b. “Prices versus Quantities versus Bankable Quantities.” *Resource and Energy Economics* 34(4), pp. 607-623.

[7] Fell, Harrison and Richard D. Morgenstern. 2010. “Alternative Approaches to Cost Containment in a Cap-and-Trade System.” *Environmental and Resource Economics* 47(2), pp. 275-297.

[8] Flood, Robert, Nancy Marion, and Juan Yepez. 2012. “A Common Framework for Thinking about Currency Crises,” In Lucio Sarno, Jessica James, and Ian Marsh (eds.), *Handbook of Exchange Rates*, John Wiley & Sons, Inc., Chapter 25.

[9] Gaudet, Gerard, Michel Moreaux and Stephen W. Salant. 2002. “Private Storage of Common Property.” *Journal of Environmental Economics and Management* 43(2), pp. 280-302.

[10] Gaudet, Gerard, and Stephen W. Salant. 2002. “The Effects of Periodic Quotas Limiting the Stock of Imports of Durables ” *Journal of Economic Theory* 109(2), pp. 402-419.

[11] Hasegawa, Makoto and Stephen Salant. 2012. “Cap-and-Trade Programs under Continual Compliance.” Discussion paper 12-33. Washington, DC: Resources for the Future.

[12] Holland, Stephen P. and Michael R. Moore. 2012. "Market Design in Cap and Trade Programs: Permit Validity and Compliance Timing." *Journal of Environmental Economics and Management* 66(3), pp. 671-687.

[13] Jacoby, Henry D. and A. Denny Ellerman. 2004. "The Safety Valve and Climate Policy." *Energy Policy* 32(4), pp. 481-491.

[14] Kling, Catherine and Jonathan Rubin. 1997. "Bankable Permits for the Control of Environmental Pollution." *Journal of Public Economics* 64(1), pp. 101-115.

[15] Krugman, Paul. 1999. "Currency Crises," available at: http://web.mit.edu/krugman/www/crises.html.

[16] Kydland, Finn E. and Edward C. Prescott. 1977. "Rules Rather than Discretion: The Inconsistency of Optimal Plans." *Journal of Political Economy* 85(3), pp. 473-492.

[17] Leiby, Paul and Jonathan Rubin. 2001. "Intertemporal Permit Trading for the Control of Greenhouse Gas Emissions." *Environmental and Resource Economics* 19(3), pp. 229-256.

[18] McWilliams, Michael, Michael R. Moore, Gloria Helfand, and Yimin Liu. 2011. "Explaining Price Movements of the SO_2 Market," work in progress.

[19] Potomac Economics. 2014. "Market Monitor Report for Auction 23." available at http://www.rggi.org/docs/Auctions/23/Auction_23_Market_Monitor_Report.pdf.

[20] Rubin, Jonathan D. 1996. "A Model of Intertemporal Emission Trading, Banking, and Borrowing." *Journal of Environmental Economics and Management* 31(3), pp. 269-286.

[21] Salant, Stephen W. 1983. "The Vulnerability of Price Stabilization Schemes to Speculative." *Journal of Political Economy* 91(1), pp. 1-38.

[22] Salant, Stephen W. and Dale Henderson. 1978. "Market Anticipations of Government Policies and the Price of Gold." *Journal of Political Economy* 86(4), pp. 627-648.

[23] Samuelson, Paul A. 1957. "Intertemporal Price Equilibrium: A Prologue to the Theory of Speculation." *Weltwirtschaftliches Archiv* Bd. 79, pp. 181-221.

[24] Samuelson, Paul A. 1971. "Stochastic Speculative Price." *Proceedings of the National Academy of Sciences* 68(2), pp. 335-337.

[25] Stavins, Robert N. 2012. "Low Prices a Problem? Making Sense of Misleading Talk about Cap-and-Trade in Europe and the USA." available at http://www.robertstavinsblog.org/2012/.

[26] Stocking, Andrew. 2012. "Unintended Consequences of Price Controls: An Application to Allowance Markets." *Journal of Environmental Economics and Management* 63(1), pp. 120-136.