What Ails the European Union's Emissions Trading System? Two Diagnoses Calling for Different Treatments

[1]

May, 2015

Abstract

In theory, how a fixed number of storable pollution permits are allocated in a cap-and-trade program should not affect intertemporal prices unless participants fail to receive permit endowments before they plan to use them. "Backloading" can create inefficiency; "frontloading" cannot. The European Union's Emissions Trading System, however, is regarded as a counterexample where frontloading itself is creating inefficiency. This view underlies current policy proposals to backload permits or to create a Market Stability Reserve. The goal of these policies is to shrink the current inventory of permits carried by the private sector without tightening the cap. We question the most prominent theory of why frontloading has been excessive by comparing its implications to a theory that attributes recent movements in the spot price of permits to ongoing regulatory risk of a price collapse much like what occurred in the 1970's in anticipation of the devaluation of the Mexican peso or the sale of massive government gold stockpiles. Correct diagnosis should precede treatment advice: if frontloading is excessive, inefficiency can be eliminated by suitable backloading of permits; if regulatory risk is excessive, however, backloading either directly or with a market stability reserve is unlikely to reduce inefficiency.

Keywords: cap and trade, emissions trading, market stability reserve, peso problem, regulatory uncertainty

[1]I would like to thank Dallas Burtraw, Harrison Fell, Karsten Neuhoff, Richard Sandor, and Luca Taschini for valuable discussions leading up to this preliminary draft. Hang Yin provided valuable assistance with the diagrams. The research was supported by RFF's Center for Energy and Climate Economics, Mistra's Indigo research program, Stiftung Mercator, and DIW Berlin.

1 Introduction

Cap-and-trade programs appear to be the politically preferred way to curb carbon emissions. According to received theory (Hasegawa and Salant, 2015) as long as the government makes permits available before they are needed to implement the efficient program, emissions trading with bankable permits will induce the regulated sector of profit-maximizing firms to abate cumulative emissions by the requisite amount in the manner that costs society the least. One way that this necessary condition of permit availability can be satisfied is to award all the permits required for the entire planning horizon at the outset rather than through auctions over the entire horizon. With a planning horizon extending to 2050, the requisite initial "bank" of permits would be huge. According to received theory, however, "frontloading" a fixed number of permits in this way would cause no inefficiencies; inefficiencies arise only if permits are backloaded to such an extent that implementing the least-cost program becomes infeasible.

For reasons we will investigate, the European trading system does not appear headed toward achieving its cap in the least-cost manner. The current permit price is extremely low and, it is claimed, this is not merely the result of overallocating permits and of unexpected low permit demand due to the financial crisis and its aftermath.[2] Given the current bank and the scheduled auctions, the current price should be closer to 20 euros per tonne rather than to the current 5 euros per tonne—again, assuming the sequence of future marginal abatement costs curves has been approximated correctly.

Prices matter because they influence behavior. A low permit price today makes mitigation unattractive. As *The Economist* (2013) lamented: "Cheaper carbon makes heavily polluting coal cheaper than cleaner gas." Low prices also stifle investment and cost-reducing technical change. Finally, if intertemporal smoothing of abatement costs requires a 20-euro current price, then much lower prices in the near term must be accompanied by much higher prices later on since the same cap limits cumulative emissions on both paths.. Hence, given the current price trajectory, the induced path of abatement will achieve the cap in an unnecessarily expensive way.

If there is an inefficiency, what could explain it? Several theories have been suggested. All can explain why the current permit price is low but rising faster than the riskless rate of interest. Although all predict that the required abatement will not occur at least cost and that investment and technical change will be depressed in the short run, they differ as to which remedy is appropriate. Proper treatment of the ailing EU ETS requires correct diagnosis of its ailment.

One diagnosis (Neuhoff et al., 2014) posits that regulated firms are willing to carry up to some maximum number of permits at the riskless rate of return but no more at any rate. If more permits are put into the system, speculators must hold them and—according to this theory—speculators typically require a much higher rate of appreciation to compensate them for their exposure to inter-period price volatility. If this diagnosis is correct, the inefficiency would disappear if policy makers could somehow eliminate the demand shocks that cause the price volatility or, more realistically, if they injected permits into the system slowly enough that the regulated firms them-

[2] Harrison Fell conjectures that even without any inefficiency, a severe downward demand shock, coupled with the persistence assumed in his model and a 3% interest rate, could at least partially account for the low prices we observe in the market; but he has not yet had time to assess this possibility quantitatively.

selves would be comfortable holding the entire inventory, thereby eliminating the need to attract speculators.[3]

A second diagnosis (Salant and Henderson, 1978) posits that market participants fear that, at a date they cannot predict, a new regulatory policy will be announced that will dramatically affect permit prices. This policy may result in an upward or a downward jump in prices. However, as long as the expected value of the price following the announcement is lower than the price prevailing before the announcement, (1) the market price in anticipation of that announcement will be depressed and (2) the rate of increase of the price must be faster than the riskless rate of return to entice risk-neutral agents to hold permits despite the downside regulatory risk. If this diagnosis is correct, the appropriate treatment would be to design policy to minimize regulatory uncertainty. Neither remedy would be effective if the other diagnosis were correct.

If either diagnosis is correct, it must also explain the precipitous jumps in price that have occurred since 2011. Among the most pronounced jumps noted by Reuters (February 22, 2013) were the following:

- June 22, 2011—the European Commission published its draft Energy Efficiency Directive, prompting a deep sell-off on the EU ETS from which it has never really recovered
- December 20, 2011—the European Parliament's Environmental Committee voted by a margin of one vote in favor of removing 1.4 billion permits and by a wider majority to take away a "significant number" of permits. EU carbon prices rose as much as 40%
- January 24, 2013—European Parliament's industry committee in a non-binding opinion rejected the idea of backloading, sending prices down 60%

This Reuters chronology was published in February 2013. Prices had fallen from 20 euros per ton in 2011 to 5 euros per ton in early 2013. But the process is continuing and so we should update it:

- April 16, 2013—the European Parliament narrowly rejected by a vote of 334 votes to 315 the backloading proposal to delay the sale of some permits, sending the price to a new low of 2.75 euros per ton on April 17, a drop of 40%

In interpreting the market reaction to the April 16 vote, *The Economist* led with the headline "Carbon trading: ETS, RIP?" noting that some environmentalists feared "the whole edifice of EU climate policy is about to crumble." Bloomberg News led with the headline "EU Carbon Permits 'Worthless' without Change of Rules, UBS Says." The secretary-general of an electricity providers' association likened the permits to "junk bonds," while *Time* tried to look beyond the wreckage to a brighter future, noting "That might leave the door open for other policies, including a straight carbon tax,

[3] A closely related diagnosis (Bredin and Parsons, 2014) postulates an "inconvenience yield" (a convenience yield net of transport and all other carrying costs that is negative). This hypothesis is based on systematic patterns in the data on spot and futures prices but, given the absence of costs of storing permits, the source of the inconvenience has yet to be identified. The implications of this hypothesis, however, are clear. Private agents would hold permits only if the spot price was anticipated to appreciate enough to compensate them for the inconvenience. If this diagnosis is correct, large inventories of permits would result in the cap being achieved at an unnecessarily high social cost but, in contrast to the theory of Neuhoff et al., elimination of demand shocks would not diminish the inefficiency.

more support for renewables or increases in R&D funding for carbon-free power. We could use all three, but carbon markets may be *finished*" [my emphasis].

Reuters interpreted these price jumps as reactions to emerging information that changed the market's probability assessment of what regulatory intervention might occur. Whatever their cause, presumably none of these price jumps was widely anticipated. For, if an upward price jump was widely anticipated, everyone would have tried to buy permits the moment beforehand and there would have been no one willing to sell them. As for an anticipated price drop, agents holding permits would have tried to sell them immediately beforehand, knowing that they could be repurchased at smaller expense a moment later; but, immediately before the drop, there would have been no one willing to purchase these permits. Indeed, according to *Time*, the April 16 vote was a development that "surprised observers."

To summarize, a credible explanation of what has been happening in the spot market for carbon permits must explain not only the low, rapidly rising price but also these price jumps in both directions. Finally, if any of these diagnoses is correct, its implications for futures prices must be consistent with what has been observed in the futures market. Diagnosing the ailment correctly is important since appropriate treatment depends on correct diagnosis.

We proceed as follows. In section two, we characterize the abatement path that achieves the cap at least cost. In section three, we explain why, according to standard theory, this path results under a cap-and-trade program provided the availability condition is satisfied. In section four, we discuss two explanations for the observed behavior of the spot price and whether they can be distinguished empirically using information from the futures market. Section five concludes.

2 The Least-Cost Benchmark

Let q_t^i denote firm i's emissions at time t and $\bar{q}_t^i (\geq q_t^i)$ denote its emissions in the absence of any abatement. Let a_t^i denote firm i's abatement in year t. That is, $a_t^i = \bar{q}_t^i - q_t^i \geq 0$. We assume that firms differ in their cost of abating. Denote by $C_t^i(a_t^i)$ firm i's *total* cost of abating a_t tons in year t. Although costs of abating differ for each firm, they share certain common features. It is costless for each firm to undertake no abatement ($C_t^i(0) = 0$), more expensive for a firm to undertake an additional ton of abatement, and this additional cost of increased abatement is higher the more abatement is undertaken. That is, $C_t^i(a_t^i)$ is strictly increasing and strictly convex for $a_t^i > 0$. Denote the *marginal* cost of abatement by firm i in year t as $c_t^i(a_t^i)$. Since in the absence of regulation, firms would have expanded emissions until costs no longer declined, we assume that $c_t^i(0) = 0$. Finally, we assume that the marginal cost of eliminating all emissions at any firm in any year is prohibitive: as $a_t^i \to \bar{q}_t^i, c_t^i(a_t^i) \to \infty$.

Suppose one wishes to reduce aggregate emissions of firms over a T-year horizon to a specified goal $\bar{G}$. Since any path of emissions satisfying $\sum_{t=1}^T \sum_{i=1}^N q_t^i = \bar{G}$ achieves this goal, it can be achieved in an infinite number of ways. However, only *one* way minimizes the present value of aggregate abatement costs among firms and across time.[4]

[4]The least cost program solves the following problem: $\min_{q_t^i \in [0, \bar{q}_t^i]} \sum_{t=1}^T (1+r)^{(1-t)} \left[\sum_{i=1}^N \int_{z=0}^{\bar{q}_t^i - q_t^i} c_t^i(z) dz \right]$ subject to (1) $a_t = \sum_{i=1}^N \bar{q}_t^i - q_t^i$ and (2) $\sum_{t=1}^T a_t = \sum_{t=1}^T \bar{q}_t^i - \bar{G}$. Since the minimand is strictly convex and the constraints are linear, the solution is unique. To find it we minimize, for each year, the sum in square

What are the characteristics of this solution?

It should be clear that any way of satisfying the goal that results in any two firms having different marginal costs of abatement in any of the T years cannot minimize aggregate abatement costs. For, in that case, it would be cheaper for the firm with the lower marginal abatement cost in that particular year to undertake a ton more abatement and the firm with the higher marginal abatement cost then to undertake a ton less abatement. This would achieve the same emissions goal $\bar{G}$ at a lower cost. It follows that a *necessary* condition for the goal to be achieved at least cost is that, in any one year, all firms have equal marginal costs of abatement. Denote the lowest cost of achieving aggregate abatement $a_t = \sum_{i=1}^{N} a_t^i$ in year t as $C_t(a_t)$ and its derivative as $c_t(a_t)$. Since a one-ton increase in aggregate abatement in year t can be achieved by increasing the abatement by one ton at any of the firms, $c_t(a_t)$ also turns out to equal this common marginal cost of all the firms in year t.

There are, of course, many ways to achieve the specified emissions goal $\bar{G}$ over the T years where, in any one year, all firms have equalized their marginal costs of abatement. For example, firms can do no abatement in the first half of the horizon (resulting in a common marginal cost of zero) and abate enough in the remaining years, while equalizing their marginal costs in any one year, to achieve the goal. Or firms can do no abatement in odd numbered years but sufficient abatement in even numbered years to achieve the goal.

However, no program can achieve the specified emissions reduction at least discounted cost unless the common marginal cost of abatement in each year has the same present value. For suppose the common marginal cost in year t had a strictly higher present value than the common marginal cost in year $\hat{t}$. Then if any firm in year t reduced its abatement one ton and any firm in year $\hat{t}$ increased its abatement by one ton, the present value of aggregate abatement costs would decline and yet the goal of reducing emissions to $\bar{G}$ would still be achieved.

In short, the cost-minimizing program (1) must reduce emissions to the goal $\bar{G}$, (2) must result in all firms equalizing their marginal costs of abatement within any one year, and (3) must result in the common marginal cost of abatement in each of the T years having the same present value. Condition (3) may be equivalently re-stated as requiring that the common marginal cost of abatement rises each year by the real interest factor $(1 + r)$. If the cost of abatement at each firm is stationary over time or is shifting uniformly downward due to technical change, then the least expensive way of achieving the emissions goal is for abatement to increase each year. If baseline emissons would be steady, this in turn means that emissions in the least cost program are largest at the outset and decline annually.

The information requirements of this least-cost solution seem daunting. After all, no one except the manager of a regulated firm, can be expected to know the cost of reducing that firm's emissions. No central planner or bureaucrat in Washington, Sacramento, or Brussels could possibly know this about every regulated firm. Yet the least-cost solution *can* be implemented: the genius of market mechanisms (emissions taxes, cap-and-trade, etc.) lies in delegating decisions about abatement at each firm to the firms themselves.

---

brackets subject to constraint (1), define the minimized value as $C_t(a_t)$, substitute it for the inner sum, and minimize the outer sum subject to constraint (2).

3 The Standard Theory of Cap and Trade

A pure cap-and-trade program is one way to reduce emissions to $\bar{G}$. The government initially issues $\bar{G}$ permits and requires that one be surrendered for each ton of carbon dioxide emitted. At time t, the permits will be worth some market-determined price, p_t. If, in complying with the regulations, a firm abated so much that its marginal cost of abatement strictly exceeded the permit price, it would not be behaving in its own self-interest. For, it could reduce its compliance costs by decreasing its abatement by a ton and instead purchasing a permit to cover that ton of emissions. Similarly, if in complying, a firm abated so little that its marginal cost of abatement were smaller than the price of a permit, it would not be minimizing costs. For, it could reduce its compliance costs by abating another ton of emissions and selling at a price higher than the resulting additional abatement cost the permit that would have been used to cover that ton. It follows that, at an optimum, every firm will set its marginal cost of abatement equal to the permit price. The permit price in year t would adjust so that the number of permits firms want to sell then equals the number of permits firms wish to buy. Since every firm, whether a buyer or seller of permits, sets its marginal cost of abatement equal to the permit price (and hence to each other), aggregate abatement in year t is achieved at least cost.

For the *intertemporal* necessary condition to be solved, permits must be bankable. As we will see, however, even bankable permits do not guarantee that the market will achieve the emissions target at least cost. The competitive equilibrium in the permit market depends on the aggregate initial bank (B_1) and the amount auctioned ($\bar{y}_t$) in every year. Let x_t^i denote firm i's year-t's purchase of permits (if positive) or sale (if negative). To determine the dynamic equilibrium under cap-and-trade, we first consider the problem of each price-taking firm: given any sequence of permit prices $\{p_1, \ldots, p_T\}$, firm i chooses $\{q_t^i, x_t^i\}_{t=1}^T$ to minimize

$$\sum_{t=1}^{T}(1+r)^{(1-t)}\{p_t x_t^i + \int_{z=0}^{\bar{q}_t^i - q_t^i} c_t^i(z)dz\} \tag{1}$$

subject to:

$$B_1^i = \bar{B}^i \tag{2}$$

$$B_{t+1} = B_t^i + x_t^i - q_t^i. \tag{3}$$

The model is closed by setting p_t for $t = 1, \ldots, T$ so that in every year $\sum_{i=1}^N x_t^i = \bar{y}_t$.

At the optimum, every firm i ($i = 1, \ldots, N$) sets its marginal cost of abatement in each year to p_t whether it purchases additional permits from the market ($x_t^i > 0$) or sells them to the market ($x_t^i < 0$). In addition, if it carries a strictly positive bank into year $t+1$ ($B_{t+1} > 0$), then $p_{t+1} = (1+r)p_t$. For, if the left-hand side were strictly larger ($p_{t+1} \geq (1+r)p_t$), infinite profits could be made by purchasing permits in year t and selling them a year later. On the other hand, if the left-hand side were strictly smaller ($p_{t+1} \leq (1+r)p_t$), it would be optimal to carry no permits ($B_{t+1} = 0$) into year $t+1$. For, if permits were needed then it would be less expensive to sell permits at t, bank the revenue for one year, and use the principal and interest to purchase the same amount of permits. This transaction can always be accomplished—with money left over as profit. Hence, if $p_{t+1} \leq (1+r)p_t, B_{t+1} = 0$.

3.1 The Availability Condition

For an equilibrium under cap-and-trade to achieve the emissions goal at least cost, enough permits must be made available before they would be needed to implement the least cost program. Otherwise, an artificial permit shortage will be created and the permit price will rise to clear the market, inducing too much abatement early in the program and too little later. In Figure 1, five functions are plotted against time. The bold curve emanating from the origin and reaching $G(T)$ at time T is the cumulative emissions ($E(t)$) in the program that achieves the goal ($G(T)$) of reduced emissions at the least discounted cost. The slope of this function at any time t ($e(t) = \dot{E}(t)$) is the flow of emissions then; the concavity of that function implies that the flow of emissions declines monotonically over time.

The other four functions are labelled $G_1(t)$ through $G_4(t)$. Each of these four functions represents a different possible path for the cumulative allocation of permits. All allocate $G(T)$ permits cumulatively by time T. The slope of each function at any time t ($g_i(t) = \dot{G}_i(t)$) is the flow of permits allocated at time t. Since permits are often allocated in a finite sequence of auctions, we have drawn $G_3(t)$ as a step function. Since $G_i(t) \geq E(t)$, it is feasible in each case for the private sector to implement the efficient solution. Hence, each of these four alternative allocation paths would generate the same path of permit prices rising at the riskless rate of interest and reducing cumulative emissions to $G(T)$ in the least costly manner.

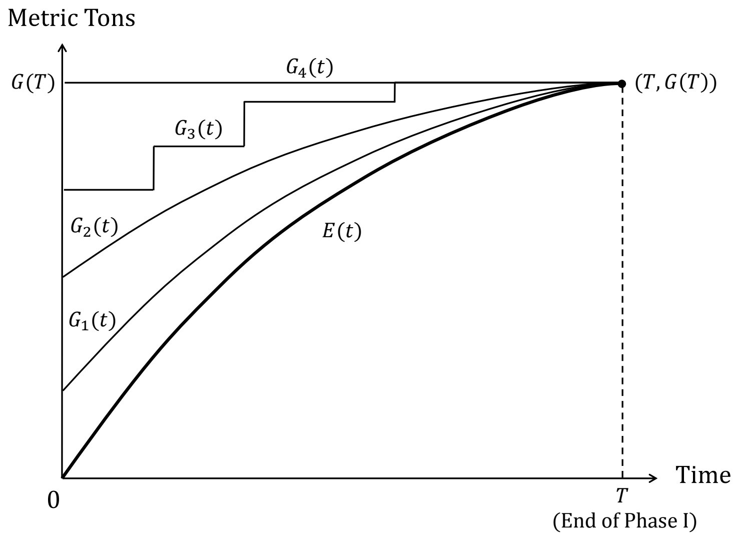
Any allocation path that is uniformly weakly smaller than another path "backloads" it. That is, the same number of permits is eventually distributed on both paths but, in the first part of the planning horizon, the cumulative release of permits is smaller on the lower path. Conversely, any allocation that is uniformly weakly larger than another path but provides the same number of permits by the end of the planning horizon "frontloads" it.

The distance between a given cumulative allocation at time t and the cumulative emissions at time t is the bank ($B_i(t)$) at time $t : B_i(t) = G_i(t) - E(t)$. In $G_4(t)$, the permits corresponding to the entire cap are allocated at the outset and the bank is huge. In $G_1(t)$, on the other hand, the bank is initially small and diminishes over time. In the standard theory, the same abatement path would be induced by any allocation path provided it never falls below $E(t)$. $G_i(t) \geq E(t)$ is a necessary condition for efficiency. Hence, the diagram also illustrates cases where backloading has no effect on prices.

If the availability condition is violated, however, inefficiency must occur. To illustrate, let B denote the number of permits allocated at the outset and assume that an additional $\bar{y}$ per unit time are auctioned throughout the planning horizon where $B + T\bar{y} = G(T)$. In terms of Figure 1, such allocation paths would be lines through the same point $(T, G(T))$ but with different vertical intercepts. Some lines (like $G_4(t)$) would not violate the availability condition. Others, like the ray from the origin to point $(T, G(T))$ would cut $E(t)$, violating the availability condition and hence creating inefficiency..

To illustrate the price consequences of such inefficient allocations, suppose that the aggregate baseline emission is stationary and denote it $\bar{q}$. Assume that the common marginal cost of emissions in each period is stationary and denote it as $c(\cdot)$. If nothing is auctioned annually (implying that $B = G(T)$), then the equilibrium price path would rise throughout by the interest factor and the emissions target would be achieved at least cost. At the other extreme, suppose $B = 0$ (implying that $G(T)/T$ is auctioned

Figure 1: Four Permit Allocations That Satisfy the Availability Condition

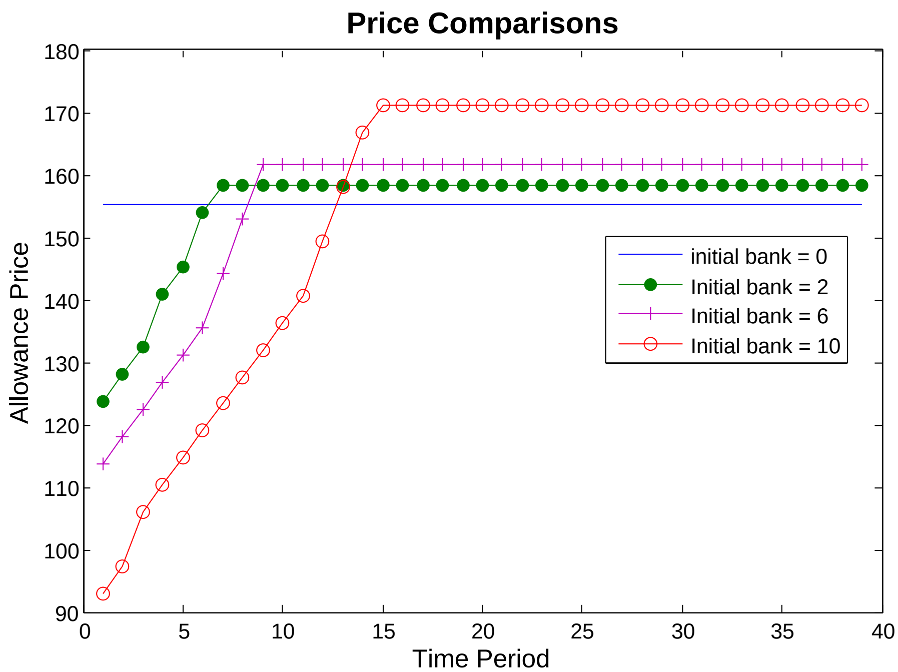


annually). An artificial permit shortage would be created and firms would have to use every permit that is auctioned to cover contemporaneous emissions and would bank no permits for future use. The equilibrium permit price would then be $p_t = c(\bar{q} - \bar{y})$ for $t = 1, \ldots, T$. Between these two extremes lie intermediate cases. Figure 2 depicts the *family* of equilibrium price paths that would be generated if the initial bank were reduced and the stationary size of the auction increased to achieve the same emissions target.

As the bank is reduced and the auction size is increased to maintain the cap, the resulting equilibrium price paths coincide at first with the benchmark path. When the initial bank gets low enough, however, an artificial permit shortage is created and the equilibrium price path starts higher, rising at the rate of interest during an initial phase ending at $\tilde{t}$. After that, no permits are carried from one year to the next and the permit price remains constant at $\tilde{p} = c(\bar{q} - \bar{y})$, crossing the benchmark path as it must if cumulative abatement over the T years is the same on the two paths. If the bank is reduced further (while raising the auction size to maintain the cap), the equilibrium price path will start even higher, its rising portion will end sooner ($\tilde{t}$ will decrease) resulting in the same cumulative abatement on each of these paths, and the price will then remain constant at a lower level ($\tilde{p}$ will decrease) since no permits are banked after the price ceases to rise and the larger auctions result in less abatement and a lower common marginal cost across firms. In the extreme case of $B = 0$, the initial phase disappears entirely ($\tilde{t} = 0$) and the common marginal cost across firms is $c\left(\bar{q} - G(T)/T\right)$. Each allocation path that starts with a smaller bank but injects permits at a faster rate so that $G(T)$ permits continue to be injected by the end of the planning horizon backloads the uniformly higher allocation paths. Hence, the diagram depicts cases where backloading *can* raise prices and introduce inefficiency.

"Good theory," Nobel Laureate Robert Solow (1974, p. 10) once observed, "is usually trying to tell you something, even if it is not the literal truth." In this context, it is saying that for cap-and-trade to be efficient, permits must be sufficiently frontloaded.

Figure 2: Shrinking the Initial Bank Too Much Violates the Availability Condition and Causes the Cap To Be Achieved at Excessive Social Cost



Attempting to achieve a cap of a given size while backloading the bank risks introducing inefficiency.

The analysis also implies that, if one wishes to tighten the cap, one should do it from the *far end.* If one can credibly persuade the market that the size of later auctions will be reduced (possibly to zero), market participants will *foresee* that they will have insufficient permits and will attempt to purchase more of them at the outset. This will drive up the current price of permits and will induce increased current abatement. The permits the firms save can be banked and used when government curtails its auctions.

Suppose the government initially wished to reduce cumulative emissions to G_0 over the planning horizon and made permits available by grandfathering $G_0(0)$ permits and then auctioning the rest continually at a constant rate.[5] Assume that this allocation path satisfied the availability condition which is necessary for efficiency as illustrated in Figure 3. If the government wanted to tighten the cap to $G_i < G_0$, for $i = 1, \ldots 3$, it could do so by grandfathering an unchanged number of permits and then auctioning additional permits at an unchanged rate but terminating the auctions the moment the cumulative number of permits issued reaches the new, tighter cap. In Figure 3, we show that these three tighter allocation paths must also satisfy the availability condition and hence would achieve the respective caps at least cost.

The price paths that are induced by these government policies are depicted in Figure 4. Each of them rises by the riskless rate of interest; but higher paths induce higher abatement throughout the planning horizon and hence smaller cumulative emissions.

[5] Grandfathering no permits ($B = 0$) but frontloading them by auctioning G_0 in the first auction would produce the same results.

Figure 3: Tightening the Cap While Continuing to Satisfy the Availability Condition

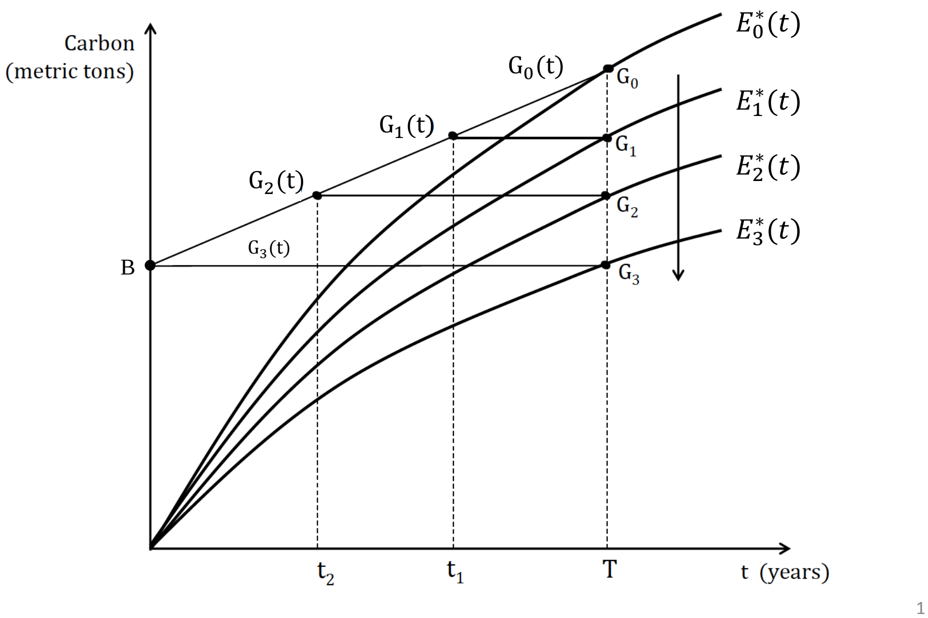
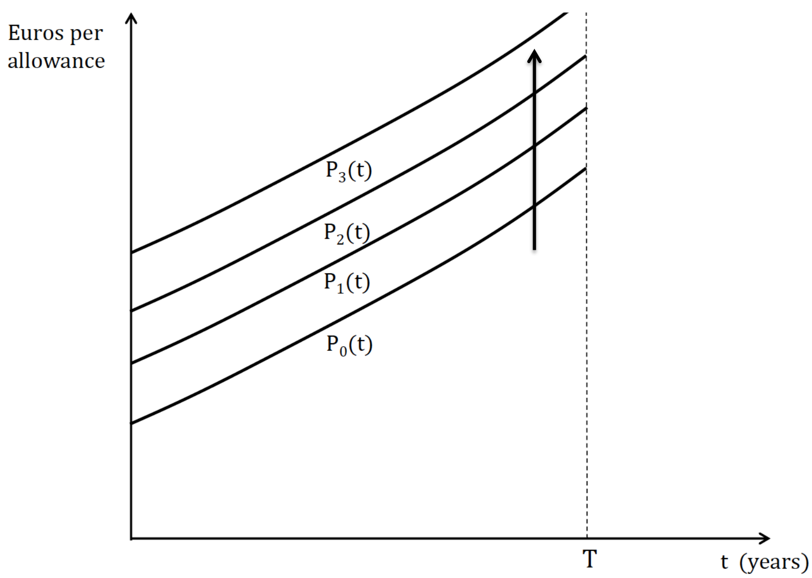


Figure 4: Efficient Price Paths for Successively Tighter Caps



4 Two Diagnoses for What Ails EU ETS

The recent experience with cap and trade in the European Union raises questions about whether the standard theory summarized in section three requires revision. In what follows, we consider two theories of what ails EU ETS. We will simplify by treating the two theories separately. However, the reader should recognize that the explanations may overlap. For example, the *reason* that speculators in this market may require a rate of return as high as 15% to hold permits in the theory of Neuhoff et al. (2014) may be the downside regulatory risk described in the Salant-Henderson theory (1978).

4.1 Diagnosis 1: The Hedging Corridor (Neuhoff et al.)

According to Neuhoff et al. (2014, p. 15-16), "Interviews suggest that the power sector's capacity to bank beyond the need to hedge future emissions was constrained by risk management requirements. Additional surplus allowances could only be banked as speculative investment. This might have required a new type of investor willing to carry the carbon price risk. If these investors required higher rates of return in order to bank allowances, the current price *had to decline* [my emphasis] to a level that allowed for such returns in subsequent years. Higher discount rates required by speculative investors might have further depressed current prices. The growing gap between excess supply and hedging volumes coincides with the significant price decline from around 15 to an average of 7 Euro/tonne of CO_2 in 2012...However, our quantitative analysis based on annual reporting is not sufficiently precise to identify the exact point at which surplus allowances in the market exceeded hedging volumes."

In my view, there are two weaknesses with this explanation. First, even if there is a well-defined "hedging corridor" beyond which hedgers will not venture, the price would not jump in either direction when the bank strayed outside that corridor; if the required rate of return jumped from 5% to 15%, the price path would be continuous but would cease to rise at 5% and would begin to rise at 15%. As explained in the introduction, for a jump in the spot price (as opposed to its time derivative) to occur in equilibrium, *some unanticipated* event must occur.

Thus, the theory of Neuhoff et al. provides no explanation for the various upward and downward price jumps that have occurred. That more offsets were used than had been predicted could explain one downward jump. Most of the others seem related to announcements that either raised or lowered the chance that particular policies would be implemented, a phenomenon that is not part of the theory of Neuhoff et al. (2014).

In addressing the sharp drop in prices following the April 16 vote, Neuhoff dismisses alternative explanations of the April 17 price drop (personal correspondence, December 5, 2014): "Alternative interpretations of the channel through which the vote of the European Parliament could have influenced the allowance price are in principle possible, but are unlikely to alone explain the large price drop. First of all, the vote could have been interpreted by market participants as a reflection of the overall support for EU ETS. The failure of the European Parliament to support back-loading could, therefore, also be an indication that it will become more difficult to determine tight emission caps in future votes, and this would translate into lower expectations of post 2020 allowance prices. However, since the vote was very tight, it seems unlikely that the result would have been interpreted so keenly."

Recall that the popular press treated this vote as almost surely signifying the death of EU ETS—because, if such a weak proposal could not secure enough votes, then any more profound reform of EU ETS seemed doomed. Reflecting on the significance of the April 16 vote, *The Economist* (April 20, 2013) concluded: "More profound reforms—even assuming they could be negotiated—would take years because they would have to be approved by all 27 EU governments."

The second weakness in the theory is its failure to explain why the "risk management requirements" that constrain hedgers would not *change* in the face of extremely attractive capital gains.

Despite these weaknesses, it is nonetheless worthwhile to deduce the implications of this theory for futures prices. According to Neuhoff (personal correspondence, December 2, 2014), the theory predicts that the futures price now quoted for delivery x months in the future should be smaller than the spot price currently expected to prevail then due to a risk premium. Robert Pindyck (2001) makes a similar claim in his discussion of oil prices.

The following is my own understanding of these claims. Suppose everyone is identical in their beliefs, utility (scaling) function, and endowments. Suppose there is a single consumption good. Suppose there are two states of nature that may be realized in x months—a high spot price of permits and a low spot price. Suppose that when the permit price is high, the exogenous endowment of the consumption good is also high and when the permit price is low, the exogenous endowment of the consumption good is low. Suppose the high spot permit price in x months is given and the lower permit price then is given. An agent can consume his risky endowment or can alter his state-contingent consumption by buying or selling futures contracts. If he goes long, he augments his endowment when the spot price is high but reduces it when the spot price is low; if he sells short, he augments his endowment when the spot price is low but reduces it when the spot price is high.

More formally, each agent chooses to buy q futures contracts (if $q < 0$ he would be selling them) to maximize his expected utility of consumption:

$$\max_{c_H \geq 0, c_L \geq 0, \text{ and } q} \quad \pi_H U(c_H) \quad + \quad \pi_L U(c_L) \text{ subject to} \tag{4}$$

$$c_i \quad = \quad \bar{\omega}_i + (p_i - F)q \text{ for } i = H, L. \tag{5}$$

If the agent takes no position in the futures market ($q = 0$) then his state-contingent consumption equals his state-contingent endowment. If, however, he buys $q > 0$ futures contracts, he can move northwest along the budget line through $\bar{\omega}$ in Figure 5; he can move in the opposite direction by selling contracts short. It is straightforward to verify that the slope of the budget line is $-\frac{dc_H}{dc_L} = \frac{p_H - F}{-(p_L - F)} > 0$.

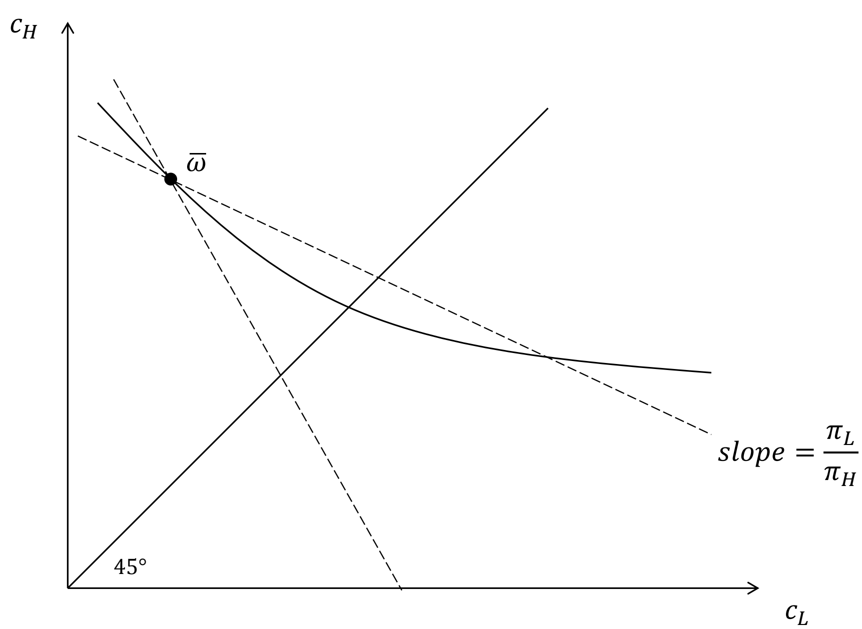
What relation must the futures price have to the expected spot price in the competitive equilibrium? In order for agents to be willing to consume their risky endowments (as they must in the competitive equilibrium if *everyone* is identical), the futures price must fall below the expected spot price. That is, since the magnitude of the slope of each indifference curve on the 45^0 line in Figure 5 is π_L/π_H, any tangency that occurs above the 45^0 line will have:

$$\frac{p_H - F}{-(p_L - F)} > \frac{\pi_L}{\pi_H}, \tag{6}$$

which in turn implies that the futures price is strictly below the expected spot price.

Intuitively, if this were not the case, agents would try to short the futures contract in an attempt to smooth their consumption across states by augmenting their endowment

Figure 5: Futures Price of a Permit Is Below the Expected Spot Price of a Permit If Agents Are Risk Averse and State-Contingent Endowments Are Positively Correlated with Permit Prices



when it would be low (in the low-price state) and reducing their endowment when it would be high (in the high-price state).

Note that futures price to be strictly smaller than the expected spot price, there must be not only (1) risk aversion but also (2) a state-dependent endowment that is larger when permit prices are high than when they are low ($\omega_H > \omega_L$). Note also that the futures price would be below the expected spot price under these conditions even if everyone were identical (even if there were no distinction between hedgers and speculators) and no inventory of permits were carried.

Now in the context of the theory of Neuhoff et al., there *is* a distinction between hedgers and speculators. So let us assume that the identical agents referred to above are speculators. In addition, let's assume that the endowment available to each of them x months in the future is the exchange value (in terms of the consumption good ($p_i I$)) of each speculator's permit inventory plus any other sources of income (y_i), where $\omega_i = p_i I + y_i$ for $i = H, L$.

If hedgers collectively are short in the futures market, then the futures price would have to fall even further to induce the speculators to be willing collectively to hold an offsetting long position.[6]

4.2 Diagnosis 2: Regulatory Risk (Salant and Henderson)

During the last four decades, regulatory risk has been the subject of intense investigation, both theoretical and econometric, in commodity markets and foreign exchange markets whenever government intervention is an ongoing concern of asset holders. Salant and Henderson (1978) postulated that changes in the ongoing risk that the government would announce a government gold auction of a given size was responsible for movements in the spot price of gold in the 1970s (see the left panel of Figure 6).

A large literature on the "peso problem" documented similar behavior in foreign exchange markets. For more than 20 years, the spot exchange rate for pesos had been fixed at .08 dollars per peso. However, during the 1970s, the market began to anticipate that the peso would be devalued at an unknown time. This anticipation affected the price of futures contracts as the right panel of Figure 6 reflects. Before devaluation was a concern, futures contracts at t for delivery at $T > t$ traded at .08 dollars per peso just like the fixed spot rate.[7] Because of the risk that the spot exchange rate might plummet, the price of a futures contract at t for delivery at T, $F(t, T)$, was initially *below* .08 dollars per peso and rose to that level only as the time to delivery approached and the interval during which a devaluation could occur prior to settlement narrowed (see Figure 6). As Karen Lewis (2008) summarized, "Futures and forward contracts sold at a discount for much of the early 1970s. For example, the year ahead contract on June 1975 and June 1976 futures contracts sold at a discount of 2.6% and 2.7% respectively . . . Therefore the *ex post* rate of return on holding Mexican peso accounts [the gain per contract that could typically be made by buying a forward contract at t] was systematically positive. Under risk neutrality, this behavior contradicts the assumption of rational expectations since it implies that the market's forecast errors

[6] The hedgers are not depicted in the diagram since they are assumed to be constrained because of "risk management requirements."

[7] The interest rate plays no role since payment on a forward contract is made when one takes delivery although the price is locked in when one purchases the contract.

Figure 6: Regulatory Risk Affects Both Spot and Futures Prices: Left Panel Is Spot Price of Gold; Right Panel Depicts the Price of Three Futures Contracts Just Prior to Devaluation

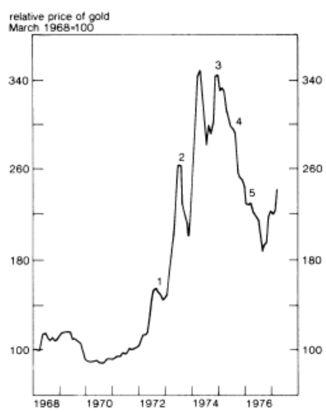
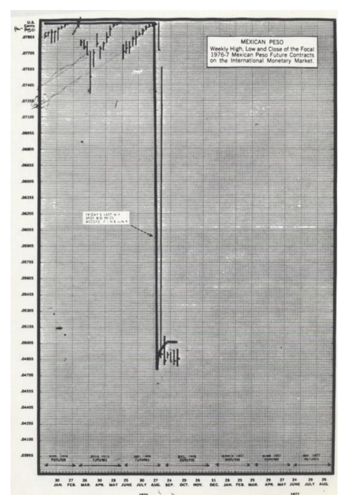


FIG. 1.—Monthly index of the dollar price of gold relative to the U.S. Consumer Price Index. *1*, First explicit reports that the United States and the IMF were considering gold sales. *2*, Reports of impending agreement to permit central bank sales. *3*, Announcement of first U.S. auction. *4*, Announcement of second U.S. auction. *5*, Agreement reached in Jamaica on IMF gold sales.



. . . were biased and serially correlated. At the end of this period, on 31 August 1976, the authorities allowed the Mexican peso to float. Subsequently, the peso fell to .06 dollars per peso, implying a *devaluation of about 46 per cent* [my emphasis]."[8]

In Salant-Henderson's analysis of the gold market and in most accounts of the peso problem, the price following the intervention is assumed to be deterministic. This does not fit the EU ETS case well. However, since agents in their model are assumed to be risk neutral, all that matters is the *expected* price immediately following the intervention.[9] Since the European Commission is considering measures that will raise the permit price, we assume that the price will jump up with some probability. On the other hand, since the popular press had all but written off EU ETS as dead after the April 16, 2013 vote, we assume that, with complementary probability, the price might jump down (possibly to zero).

Assume the price will jump down to $p^L (\geq 0)$ with probability $(1 - \lambda)$ and will jump up to $p^H \geq p^L$ with probability λ. Denote as p^A the expected price if there is a regulatory intervention.[10]

$$p^A = \lambda p^H + (1 - \lambda)p^L. \tag{7}$$

Assume the risk of a single regulatory intervention is ongoing and assume the continuous-time hazard rate for the intervention is $\alpha > 0$.[11] If there is a risk that the spot price will jump at an unknown time, then the price in anticipation of that jump must increase at the following rate:

$$\dot{p} = (\alpha + r)p(t) - \alpha p^A. \tag{8}$$

Otherwise, there would be an opportunity for arbitrage.[12] We will henceforth assume that $p^A < p(t)$. That is, although the price can jump in either direction, the mean change is downward. Equation (8) implies that, if $\alpha = 0$, the price will rise at the riskless rate (r); but in the more realistic case where $\alpha > 0$ the price will rise faster than the rate of interest given our assumption that $p^A < p(t)$.[13] Intuitively, suppose I

[8] According to Lewis, "The original source of the term [peso problem] is unknown, though some economists have attributed it to Nobel Laureate Milton Friedman." Nobel Laureate Krugman (July 15, 2008) disputes this attribution, arguing that the name "peso problem" originated in the MIT lunchroom: "Back in 1975-6 a lot of us were Rudi Dornbusch students, working on exchange rates. We were alerted to the peso issue by stuff coming from Steve Salant at the Fed; this was the same time that the Salant-Henderson model of gold markets was being drafted, and expectations—including expectations of possible sudden large price changes due to events such as gold auctions and devaluations—were very much on our minds." Considering that both authors of the classic articles on the peso problem mentioned by Lewis, Ken Rogoff (1977, 1980) and Bill Krasker (1980), were Dornbusch students (like Krugman) and that Salant-Henderson paper was annually on Dornbusch's graduate syllabus, Krugman's recollection sounds entirely plausible.

[9] This observation was made originally in a footnote of Salant and Henderson (footnote 14, p. 634) but they did not develop it; we do so here.

[10] We assume p^A and its underlying components are exogenous constants because we lack the information to endogenize them.

[11] We assume that once the first intervention occurs, no subsequent interventions are anticipated. This simplification does not affect the qualitative results we derive.

[12] The discrete-time counterpart to this differential equation is: $\frac{p_{t+1} - p_t}{p_t} = \hat{\alpha}(\frac{p_{t+1} - p^A_{t+1}}{p_t}) + \hat{r}$, where the circumflexes distinguish the discrete-time hazard and interest rate from their continuous-time counterparts.

[13] If $p^A > p(t)$ and $\alpha > 0$, equation (8) implies that $\dot{p} < rp$. This "reverse" peso problem has been discussed in the theoretical literature and documented in the empirical literature. In Chaton et al. (2009), agents in Europe are modeled as holding inventories of natural gas in anticipation that the price will jump up at a random time because of a cutoff of gas shipped through the Ukraine. In Roll (1984), orange juice futures

can sell permits today and bank the proceeds at $r\%$. If I am risk neutral and willing to hold permits instead, I must expect the same rate of return in the form of appreciation. For the mean price to rise by $r\%$ even though there is a chance that the price will fall to p^A, the price must rise faster than the rate of interest in the absence of the anticipated announcement.

What happens if—as occurred on January 24 and April 16, 2013—there is an *exogenous reduction* in the probability that a high price will occur following an announcement or a close vote? That is, what happens if λ decreases or, equivalently, if p^A decreases? It is straightforward to show that the spot price must drop and must rise at a faster rate, eventually crossing the old equilibrium price path from below. Similarly if, as occurred on December 20, 2011, there is an *exogenous increase* in the probability that a high price will occur following an announcement, then p^A rises, the spot price jumps up and must rise at a slower rate, eventually crossing the old equilibrium price path from above.

To verify this, note that equation (8) is a linear, first-order differential equation with constant coefficients. Its solution is:

$$p(t) = p(0)e^{(\alpha+r)t} - \frac{\alpha p^A}{\alpha + r}(e^{(\alpha+r)t} - 1), \tag{9}$$

where $p(0)$ is the initial price. The initial price $p(0)$ is determined by the requirement that the cumulative demand for permits over time (or, equivalently cumulative emissions) equals the exogenous cap. To show that an exogenous reduction in p^A results in an endogenous reduction in $p(0)$, assume the contrary. That is, suppose p^A decreased exogenously and $p(0)$ did not change or increased. Given the solution (9) to the differential equation, the entire price path would be uniformly higher. Since the demand for permits at any time t is a strictly decreasing function of the price at t, the cumulative demand for permits over the planning horizon would no longer be as large as the cap. Hence, $p(0)$ must fall when p^A decreases exogenously. Moreover, it cannot fall so far that the price path in the new equilibrium lies uniformly below the old path; since the cap is unchanged and the sequence of demand curves is unchanged, the two paths must cross. When they cross, the new equilibrium path must cross the old one from below since the $\dot{p}(t)$ in equation (8) will be larger. Since this must be true every time the new path crosses the old one, there must be a *unique* crossing. Prior to that time, the spot price on the new path would be strictly smaller than before; afterwards it must be larger.

Regulatory risk also affects the price of each futures contract. If agents are risk neutral, the price at t of a futures contract for delivery at $T(> t)$, must equal the expected spot price at the date of delivery. Hence,

$$F(t, T) = p(T)e^{-\alpha(T-t)} + p^A(1 - e^{-\alpha(T-t)}). \tag{10}$$

This implies, of course that at $t = T, F(t, T) = F(T, T) = p(T)$. That is, if the regulator has not intervened prior to T, the futures price will equal the spot price then prevailing.

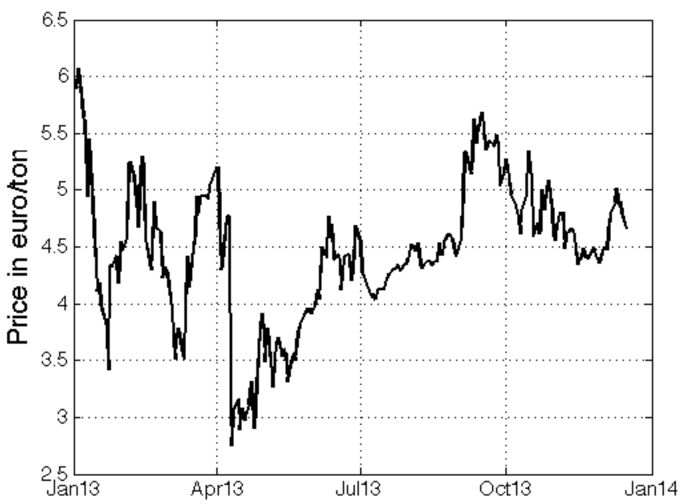
reflect the ongoing risk that a freeze will destroy the Florida crop, resulting in an upward jump in the spot price at an unknown time. The distinguishing characteristic of a peso problem is the ongoing but as yet unrealized risk of a price jump; as this last example reflects, the source of the risk may be nature rather than a policy-making body.

But prior to T, the futures price will be below $P(T)$ and will be increasing ($\dot{F} > 0$) monotonically toward the spot price that will prevail at the delivery date in the absence of a realization of the uncertainty. Differentiating, we conclude

$$\partial F/\partial t = \alpha e^{-\alpha(T-t)}(p(T) - p^A) > 0. \tag{11}$$

Hence, if the expected jump in price is downward, the futures price would be below $p(T)$ and would monotonically increase until it reached $p(T)$ at $t = T$.[14] Like the spot price, the futures price would jump down in response to an exogenous decrease in p^A.[15] The events of January 24 and April 16, 2013 had a dramatic effect on the price of the December 2014 futures contract as can be seen in Figure 7.

Figure 7: December 2014 Futures Contract as Delivery Date Is Approached: Its Price Plummeted after the January 24 and April 16 News—but Grew Rapidly Thereafter



5 Conclusion

According to standard theory, in order for a cap-and-trade program to achieve a given emissions reduction in the cheapest way, an availability condition must be satisfied. This condition is most easily satisfied by frontloading permits. If the standard theory is correct, frontloading cannot be excessive.

The European Union's Emissions Trading System, however, is regarded by many observers as a real-world example where frontloading *has been* excessive: it is claimed

[14]If $p^A > p(t)$, $\dot{F} < 0$ and the future price approaches the spot price on the date of delivery from above. Then $\dot{F} < 0$, as Roll (1984) illustrated empirically in the case of orange juice futures.

[15]I am still working out conditions sufficient for this claim to be true.

that the size of the permit inventory itself is creating inefficiency. This view has motivated several proposals (backloading and the Market Stability Reserve) intended to shrink the inventory of permits carried by the private sector without tightening the cap.

Neuhoff et al. (2014) have proposed a prominent theory for why frontloading is excessive. In their view, the power sector is only willing to carry a limited inventory of permits no matter what rate of appreciation occurs; the remainder must be carried by speculators and they require appreciation rates of at least 15%.

In this paper, we question the claim that frontloading has been excessive by comparing it to the alternate theory that attributes recent movements in the spot price to regulatory risk. The emissions trading market is likened to the peso and gold markets in the 1970's, where private agents had to take into account the significant risk that the price of the assets they were holding would plummet in value due to government intervention.

We raise questions about the theory that frontloading is excessive because we believe correct diagnosis should precede treatment recommendations. Both theories predict price paths with the same qualitative features. If frontloading is excessive, any inefficiency can be eliminated by suitable backloading of permits; if regulatory risk is excessive, however, backloading permits either directly or with a market stability reserve is unlikely to reduce inefficiency.

References

[1] Bloomberg. January 21, 2013. "EU Carbon Permits 'Worthless' Without Change of Rules, UBS Says."available at: http://www.bloomberg.com/news/print/2013-01021/eu-carbon-permits-worthless-without-change-of-rules-ubs-says.html

[2] Bredin, Don and John Parsons. 2014. "Why Is Spot Carbon So Cheap and Future Carbon So Dear? The Term Structure of Carbon Prices." Working Paper.

[3] Chaton, Corinne, Anna Creti, and Bertrand Villeneuve. 2009. "Storage and Security of Supply in the Medium Run." *Resource and Energy Economics.* 31, pp. 24-38.

[4] The Economist. April 20, 2013. "Carbon Trading: **ETS, RIP?**"available at: http://www.economist.com/node/21576388/print

[5] Hasegawa, Makoto and Stephen Salant. 2015. "The Dynamics of Pollution Permits." *Annual Reviews of Resource Economics.*

[6] Krasker, William. 1980. "The Peso Problem in Testing the Efficiency Forward Exchange Markets." *Journal of Monetary Economics* 6, pp. 269-276.

[7] Krugman, Paul. 2008. "Trivial Intellectual History Blogging,"available at: http://krugman.blogs.nytimes.com/2008/07/15/trivial-intellectual-history-blogging/

[8] Lewis, Karen. 2008. "Peso Problem." *The New Palgrave*, Second Edition.

[9] Neuhoff, Karsten, Anne Schopp, Alexander Vasa, Kateryna Stelmakh, and Rodney Boyd. 2014. "Does the Size of the Allowance Surplus under the EU ETS Matter?" Working Paper.

[10] Newbery, David. 2008. "Futures Markets, Hedging and Speculation." *The New Palgrave*, Second Edition.

[11] Pindyck, Robert. 2001. "The Dynamics of Commodity Spot and Futures Markets: A Primer. " *The Energy Journal* 22(3), pp. 1-29.

[12] Reuters. February 22, 2013. "TIMELINE—The EU's Tangled Carbon Tale,"available at: http://www.reuters.com/assets/print?aid=USL6NOBKFEM20130222

[13] Rogoff, Kenneth. 1977."Rational Expectations in the Foreign Exchange Market Revisited." Unpublished manuscript, MIT.

[14] Rogoff, Kenneth. 1980."Tests of the Martingale Model for Foreign Exchange Futures Markets."in *Essays on Expectations and Exchange Rate Volatility.* Unpublished Ph.D. dissertation. MIT.

[15] Roll, Richard. 1984. "Orange Juice and Weather." *American Economic Review* 74(5), pp. 861-880.

[16] 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.

[17] Solow, Robert. 1974. The Economics of Resources or the Resources of Economics." *American Economics Review* 64(2), pp. 2-14. May Papers and Proceedings, Richard T. Ely Lecture.

[18] *Time*. April 17, 2013."If Carbon Markets Can't Work in Europe, Can They Work Anywhere?"available at: http://science.time.com/2013/04/17-if-carbon-markets-cant-work-in-europe-can-they-work-anywhere/print/