Environmental & Resource Economics (2005) 31: 133–157
DOI 10.1007/s10640-005-1761-y

Managing Permit Markets to Stabilize Prices

RICHARD NEWELL*, WILLIAM PIZER and JIANGFENG ZHANG
*Resources for the Future, 1616 P Street, NW, Washington, DC 20036, USA; *Author for correspondence (e-mail: newell@rff.org)*

Accepted 19 July 2004

Abstract. The political economy of environmental policy favors the use of quantity-based instruments over price-based instruments (e.g., tradable permits over green taxes), at least in the United States. With cost uncertainty, however, there are clear efficiency advantages to prices in cases where the marginal damages of emissions are relatively flat, such as with greenhouse gases. The question arises, therefore, of whether one can design flexible quantity policies that mimic the behavior of price policies, namely stable permit prices and abatement costs. We explore a number of "quantity-plus" policies that replicate the behavior of a price policy through rules that adjust the effective permit cap for unexpectedly low or high costs. They do so without necessitating any monetary exchanges between the government and the regulated firms, which can be a significant political barrier to the use of price instruments.

Key words: banking, borrowing, prices, quantities, tradable permit market, uncertainty

JEL classifications: Q28, Q48, D8, L51

1. Introduction

Frequently in the course of designing new regulation, policymakers have incomplete information about the cost of compliance. Such circumstances lead to a dichotomy between otherwise equivalent price-based market mechanisms, such as taxes, and quantity-based market mechanisms, such as tradable permits. Price-based mechanisms fix the price and leave output uncertain, while quantity-based mechanisms fix the output level and leave prices uncertain. Seminal work by Weitzman (1974) establishes the conditions under which prices are preferred to quantities.

Weitzman's analysis and many that followed (Roberts and Spence 1976; Weitzman 1978; Yohe 1978) are set in a single-period, static framework. In reality, most regulations exist in a multi-period, dynamic framework, which gives rise to a feature unique to a permit mechanism over time: the potential to bank and borrow permits between periods. A limited amount of work has explored the consequences of banking (Rubin and Kling 1993; Cronshaw and Kruse 1996; Rubin 1996; Kling and Rubin 1997; Schennach 2000; Leiby and Rubin 2001), and only recently has such work explored its implications for policy instrument choice due to the effect of cost uncertainty (Williams 2001; Yates and Cronshaw 2001). This is true despite the widespread allowance of banking within permit systems.[1] Borrowing has been allowed to a much lesser extent to date, but is often raised in the context of potential climate change policy.[2]

We demonstrate that a bankable permit system in a multi-period setting can be used to create the same outcomes as a price-based system, provided the regulator can commit to certain rules. Unlike a typical permit system where the number of permits available over time is fixed, we consider a system where the number of new permits issued each period varies based on the previous period's emissions and price levels. By allowing the permit level to vary with past cost shocks, this bankable permit system exhibits the same cost flexibility as a price-based system.

This has important implications for environmental policy. When market-based instruments have been used, the political economy of environmental regulation in the United States has overwhelmingly favored tradable permit systems, with initial allocations given to existing firms, or "grandfathered." Keohane et al. (1998) suggest several explanations for this revealed preference, including that tradable permits create rents, and grandfathering distributes those rents to existing firms while also erecting barriers to entry. They also point out how direct allocation of grandfathered permits offers a degree of political control over the distributional effects of regulation, enabling the formation of majority coalitions. Taxes, on the other hand, offer none of these advantages. Rather, they transfer resources from the private sector to government and make the costs of regulation particularly visible. Plus, there is simply the stigma of being a "tax" – the "T" word.

In addition, practical and political issues arise when environmental policies require monetary exchanges with the government, as with taxes, auctioned permits, or hybrid systems involving a "safety valve" (where the government places an upper limit on permit prices through a willingness to sell extra permits at a set price). Beyond the factors mentioned above, these instruments also raise legislative and administrative difficulties because they cut across traditional institutional boundaries. Policies involving only emissions quantities may fall clearly within the historical bounds of particular environmental legislative committees and executive agencies. Policies entailing transfers to and from the federal treasury, however, may involve an entirely different set of legislative and executive actors not historically involved in environmental policymaking. In addition to the potential loss of authority for the traditional environmental policymakers, introducing new participants to the process raises real political and bureaucratic policymaking challenges that reduce the likelihood of agreement.[3]

Nonetheless, price-based policies are more efficient for many environmental problems. When uncertainty exists about the costs of abatement, and policies must be fixed before the uncertainty is resolved, price policies will lead to distinctly different outcomes than quantity policies. Pollution taxes, for example, encourage firms to reduce emissions until the marginal cost of reductions equals the tax. The tax leads to a range of possible emission levels, depending on how uncertainty is resolved, but will fix marginal cost at the tax level. Conversely, a tradable permit system will fix the level of emissions, with the permit price determined by the marginal cost of meeting the emissions constraint. The permit mechanism will therefore lead to a range of possible marginal costs, depending on how uncertainty is resolved, but will lead to a fixed level of aggregate emissions. Different expected net benefits will therefore be associated with these alternate policies.

Weitzman's (1974) insight was that, on economic efficiency grounds a flat expected marginal benefit function (relative to marginal costs) favors prices, while a steep benefit function favors quantities. Intuitively, flat marginal benefits imply a constant benefit per unit, suggesting that a tax could best correct the externality. In contrast, steep marginal benefits imply a dangerous threshold that should be avoided – a threshold that is efficiently enforced by a quantity control.

Thus, for cases where the marginal benefits of pollution control are flat relative to the marginal costs of abatement, prices are preferred on efficiency grounds. Furthermore, if marginal benefits are not only relatively flat, but are close to being constant, the price policy is not only the better of "second-best" instruments, but can actually be the first-best solution – even if there is uncertainty about costs – because it corresponds perfectly to the externality. The marginal benefits of carbon mitigation, for example, are thought to be very flat due to the stock nature of the externality, thereby strongly favoring the use of price-based instruments (Hoel and Karp 2002; Newell and Pizer 2003).[4] The same thinking applies, however, to both stock and flow externalities, the key feature being relatively flat marginal benefits of abatement rather than the decay, rate of the pollutant.

We are therefore interested in the potential of tradable permit systems incorporating banking and borrowing to mimic the behavior of a price-based regulatory system. In other words, we want to demonstrate that without actually selling permits or taxing emissions at a fixed price, a regulator can create a tradable permit program in a way that replicates the emissions consequences of an emissions tax over time. We could imagine this interest stemming from a given political objective to match a stream of observed permit prices, p_t, with a preconceived notion of the correct prices, p_t^*, regardless of where that notion comes from. Or we could imagine the motivation for mimicking a price-based system coming from a more fundamental desire to maximize expected net social benefits, which, for the case at hand, happens to lead to a preference for prices. In either case, the regulator will want to choose – and commit to – a set of rules to meet the objective of stabilizing permit prices around a particular price target or path of price targets.

In the next section, we summarize the existing literature regarding bankable permit systems and we present a simple model of firm and regulator behavior under uncertainty that allows for permit banking and borrowing. In Section 3, we demonstrate several different ways to manage a permit system so that it is equivalent to a tax on the regulated output, with some being more complex than others.

2. Permit Banking and Borrowing

2.1. PREVIOUS LITERATURE

Absent cost uncertainty and assuming competitive behavior, a system of emissions permits that allows trading, banking, and borrowing can achieve a cumulative emissions target over a fixed horizon at the least discounted cost to firms (Cronshaw and Kruse 1996; Rubin 1996). Given a constant annual permit allocation over a finite horizon and one-for-one banking and borrowing, in equilibrium firms will borrow emissions in early periods and pay them back later, with permit prices growing at the rate of discount in Hotelling fashion. This results in higher emissions in earlier periods and lower emissions in later periods. Unrestricted banking and borrowing of permits is generally not socially optimal, however, because it may increase total social damages depending on when emissions occur (Kling and Rubin 1997).

As Kling and Rubin (1997) note, the regulator can identify a permit trading ratio that is not one-for-one to induce firms to behave more in accordance with a social optimum. The permit trading ratio acts like a rate of interest, providing a return to banking and a penalty for borrowing. For flow pollutants, if marginal damages are constant and unchanging, the ideal trading ratio over time simply equals the inverse of the discount factor, e.g., one permit for $1/\beta$ permits next year, thereby exactly offsetting firms' desire to borrow emissions due to discounting. Note that in a setting with constant costs and constant allocations, firms will be indifferent to banking, borrowing, or doing neither if the trading ratio equals the inverse of the discount factor. Banking and borrowing have no value in that setting. Leiby and Rubin (2001) generalize these results to handle the case of stock pollutants and nonconstant marginal damages, finding that the optimal trading rate between periods (i.e., the trading ratio minus one) is equal to the discount rate minus the desired rate of change in permit prices.

The above studies assume full information on future abatement costs and production technology. In such a scenario, the instrument choice decision does not arise because the regulator can achieve the first-best solution through either a price or quantity policy. If there is cost uncertainty, however, Yates and Cronshaw (2001) and Williams (2001) find that outcomes differ and the choice of whether to allow banking or borrowing of permits depends, as one might expect, on the relative slopes of marginal benefits and marginal costs.[5] They find that in cases where marginal damages are less steep than marginal costs, intertemporal trading raises net benefits.[6]

Newell and Pizer (2003) take this one step further, suggesting that in the absence of persistent cost shocks, full banking and borrowing across all periods would in fact make a quantity control behave much like a price control since quantities rather than marginal costs and prices would fluctuate in response to cost shocks. They draw an analogy to how the marginal utility of consumption fluctuates only slightly in response to transient income shocks under the permanent income hypothesis. Building on that suggestion, this paper both considers the more general case with persistent shocks and addresses a practical flaw in their argument – that full banking and borrowing does nothing to preclude regulated firms from Ponzi-like borrowing to avoid compliance. In turn, this paper suggests a wide range of possible, practical policy approaches to mimic price instruments with intertemporal quantity controls.

A key element in all of the proposed policy approaches is the commitment by the regulator to a *rule* governing its behavior in the future. In many if not most situations, the rule will require the government to act in some way that is sub-optimal at the moment action is required, even though the rule enforces a broadly optimal outcome when viewed across time. This is analogous to debates in the monetary economics literature on "rules" versus "discretion" to combat unemployment and business cycles (Kydland and Prescott 1977; Barro and Gordon 1983; Haubrich and Ritter 2000). There are three relevant points from that literature. First, if policy attempts to exploit past empirical relationships this will influence the expectations of economic agents. This implies that those relationships may break down because economic agents will not persistently behave irrationally (the Lucas Critique). Second, a once-and-for-all fixed policy rule that maximizes a given objective will generally differ from a policy approach that chooses policy to re-maximize the objective every period (i.e., discretion). Third, an underlying practical question is whether the government can commit to the once-and-for- all rules in various contexts. This third question is of particular relevance in the current context; we return to it below in Section 3.

2.2 MODELING PERMIT MARKETS WITH BANKING AND BORROWING

There are two necessary elements to our policy design: (i) the mechanism governing banking and borrowing, and (ii) a rule for setting policy stringency, that is, the (aggregate) annual permit allocation. Banking and borrowing reduce – and can even eliminate – price shocks by converting them to quantity shocks, which are shifted across time through intertemporal arbitrage. The stringency-setting rule is necessary to anchor the price path at the desired level and to adapt to unexpected shocks. We entertain only design elements that do not involve money transfers between the government and the regulated firms.[7]

We consider a world with competitive markets for permits in every period where firms take prices as given. At the beginning of each period, the regulatory authority decides on the number of new permits to issue, determining supply.[8] Each individual firm chooses its emissions level and end-of-period bank of permits. The aggregate market demand for permits is determined by adding the total current period emissions to the desired bank at the end of the period, then subtracting any banked permits from the previous period.

Permits represent the right to emit a fixed amount of pollution within a particular period of time: for example, one ton of carbon dioxide in the year 2004. Banking occurs when firms present unused permits for emissions in the current year and, in exchange, get permits for the subsequent year from the regulatory authority. For each period, there exists a trading ratio (R_t) that defines the number of permits received: n permits in period t can be traded for $R_t \times n$ permits in period $t + 1$. The path of future trading ratios is known with certainty.[9] Market equilibrium in period t can therefore be written as:

$$y_t = (\bar{e}_t - a_t) - B_t + (B_{t+1}/R_t), \tag{1}$$

where y_t is the new aggregate supply of period t permits provided by the regulatory authority and the right-hand side of the equation reflects aggregate demand described above. The term $(\bar{e}_t - a_t)$ indicates the net aggregate emissions level (and use of permits) in the current period, where $\bar{e}_t$ is aggregate baseline emissions and a_t is aggregate abatement. B_t is the volume of banked permits at the beginning of period t, and B_{t+1}/R_t equals the aggregate amount of permits that firms bank at the end of period t.

Borrowing occurs when firms get n permits in the current period in exchange for the obligation to return $R_t \times n$ permits in the subsequent period. Note that the same trading ratio that applies to banking also applies to borrowing. Also, note that while the banking transaction can be completed immediately – a trade of period t permits for period $t + 1$ permits – the borrowing transaction requires some type of contract because it is not complete until the firm returns the borrowed permits, with interest, in the next period. Borrowing appears in Equation (1) as a negative value for B_t. Apart from the general rules for trading among periods, described above, the remaining elements of the permit system are an allocation or allocation rule for the supply of new permits, y_t, and specification of the trading ratios, R_t.

Before going forward, a key feature to keep in mind is the perpetual information asymmetry that we assume exists between the regulator and firm.

In particular, let $C_i\ (a_{i,t},\ \theta_t)$ represent a convex abatement cost function for each firm, i, where $a_{i,t}$ is abatement by firm i at time t and θ_t represents a mean-zero random shock to the marginal cost function that is the same for all firms and that may be correlated across time.[10,11] We assume the regulator knows the abatement cost function, $C_i\ (a_{i,t},\ \theta_t)$, but never directly observes θ_t. However, the regulator can usually infer the value of θ_t in period $t + 1$ based on the observed market price and level of abatement in period t. In contrast, the firm learns the value of θ_t at the beginning of period t and makes abatement decisions accordingly. (The firm does not know the value of future θ.) We therefore consider practical rules for setting y_{t+1} that are based on observed abatement and permit prices in period t.

2.3. INTERTEMPORAL ARBITRAGE BY FIRMS

With our competitive market assumption, individual firm i focuses on a sequence of optimal abatement ($a_{i,t}$) and banking ($B_{i,t+1}$) decisions, based on the realized cost shock (θ_t), the market permit price (p_t), the current trading ratio (R_t), and their initial banking position ($B_{i,t}$). These abatement and banking decisions also depend on future trading ratios $\{R_{t+s}\}$ and expectations about future prices $\{p_{t+s}\}$. Let $\Omega_{i,t} \equiv (p_t, \theta_t, R_{t+s|s\geq 0}, E_t\{p_{t+s|s>0}\})$ be the vector of exogenous variables for each individual firm. We should note that expectations about future prices, $E_t\{p_{t+s|s>0}\}$, inherently depend on the allocation rules $y_t(\,)$ established by the regulator. In fact, it is only because of these rules that the price path is determined. A very obvious analogy is how the U.S. Federal Reserve Board influences interest rates through a combination of setting the Federal Funds Rate and through open market operations, based in turn on preferences for supporting economic growth and avoiding inflation. With a rough idea of the Federal Reserve's rules for setting rates, firms are able to form expectations about future (forward) interest rates.

Based on this information set, firms maximize expected profits each period, which can be formulated as negative costs plus the expected discounted value of banked permits in the next period. As long as the banking/borrowing possibility exists, this suggests a recursive definition of the cost associated with the policy of the form:

$$\begin{aligned} V_t(B_{i,t}, \Omega_{i,t}) &= \max_{a_{i,t}, B_{i,t+1}} \{-C_i(a_{i,t}, \theta_t) - p_t(-a_{i,t} - B_{i,t} + (B_{i,t+1}/R_t)) \\ &\quad + \beta E_t[V_{t+1}(B_{i,t+1}, \Omega_{i,t+1})]\}, \end{aligned} \tag{2}$$

where, for each firm i and period t, $V_t(B_{i,t}, \Omega_{i,t})$ is the value of the bank (or debt) of permits at the beginning of period t, $E_t[V_{t+1}(B_{i,t+1}, \Omega_{i,t+1})]$ is the expected value of the period $t + 1$ bank of permits conditional on information known at time t (e.g., the value of θ_t), and β is a constant discount factor. Note again that i subscripts reflect firm-level variables, while the absence of an i subscript reflects an aggregate variable.[12] Maximizing the bracketed portion of Equation (2) with respect to the levels of abatement and banking to the next period yields the necessary first-order conditions:

$$-\frac{\partial C_i(a_{i,t}, \theta_t)}{\partial a_{i,t}} + p_t = 0, \tag{3}$$

and

$$-(p_t/R_t) + \beta E_t\left[\frac{\partial V_{t+1}(B_{i,t+1}, \Omega_{i,t+1})}{\partial B_{i,t+1}}\right] = 0. \tag{4}$$

That is, optimizing firms equate their marginal cost of abatement with the permit price and, in equilibrium, this price must equal the expected discounted marginal value of banking permits until the next period, after adjusting for the trading ratio. Note that application of the envelope theorem to the maximized expression (2) yields

$$\frac{\partial V_t(B_{i,t}, \Omega_{i,t})}{\partial B_{i,t}} = p_t,$$

which, when applied to (4), implies

$$p_t = \beta R_t E_t[p_{t+1}]. \tag{5}$$

In other words, banking and borrowing creates an arbitrage opportunity between periods. If expected prices do not satisfy the no-arbitrage condition given by Equation (5), an opportunity exists to make money by buying and selling permits in different periods. Profit-maximizing firms will exploit this opportunity until demand and supply in these markets re-establishes the no-arbitrage condition.

Notice that the no-arbitrage condition fixes the relationship among prices over time. Any change in the current period price affects the entire path based on Equation (5). Current period price shocks are reduced or eliminated via the no-arbitrage condition, which converts them into quantity shocks that can be moved across time through banking/borrowing.[13] Also note that (5) implies that for the regulator to achieve the desired price path, the trading ratio, R_t, should be set so that $R_t = p_t^*/(\beta p_{t+1}^*)$, as in the deterministic banking literature, with the ratio equaling the inverse of the discount factor for the case of constant prices.

We cannot stop here, however, because the no-arbitrage condition only fixes the ratio of prices between periods; it does not set the precise price level in any period. Without further policy conditions, any price path having the stipulated period-to-period price ratios, βR_t, would satisfy Equation (5), including $p_t = 0$ for all t. To complete the price-replicating quantity policy, we turn to methods for fixing the price level.[14] These methods rely on setting the overall stringency of the policy – that is, the effective number of permits in the system.

3. Managing Permit Markets to Fix the Price Path

As shown in Equation (5), in order to fix the price today at the desired level, p_t^*, we need to fix the future expected price, $E_t[p_{t+1}]$, at the desired level, p_{t+1}^*. We consider several approaches, some of which are oriented to work over a finite horizon, and others that work over an infinite horizon. Generally, a deterministic permit supply rule (e.g., $y_t = 100$) will not lead to a deterministic price path. If costs turn out to be unexpectedly and persistently high, the entire expected price path will have to rise because supply is fixed. Similarly, if costs turn out to be unexpectedly and persistently low, the market price will have to fall given a fixed supply. In order to find a permit supply rule that leaves the price path unchanged, we must turn to supply rules that depend on observed cost shocks revealed through the market.

A sufficient condition will be to either fix $E_t[p_{t+\tau}] = p_{t+\tau}^*$, $\tau \geq 1$ at some point in the future, or to impose some other constraint that rules out alternatives to the desired price path. Fixing the expected price in some future period will fix the price path in all prior periods through intertemporal arbitrage. In the case of a finite horizon problem, where banking and borrowing will end on a particular date, T, we can fix $E_t[p_T] = p_T^*$ (though the actual price in the last period will not be fixed).[15] In the case of an infinite horizon problem, we can either intermittently establish fixed expected prices by closing the bank (as in the finite horizon problem), or by carefully constructing a permit supply rule, y_t, along with finite limits on banking and borrowing. One practical issue we return to in Section 3.3 is whether the proposed rules present a commitment problem for policy makers.

Note that announcement of an explicit target price would be a valuable policy development because it would make clear to all parties what permit price outcome is intended. This would be of obvious benefit for both short-term and long-term investment planning because it would prove a clear signal of expected incremental abatement costs.

3.1 FIXING PRICES BY ADJUSTING ALLOCATIONS TO OFFSET THE BANK

For a finite horizon, or for an infinite horizon divided into discrete intervals, the regulatory authority can easily fix the expected price in the last period of the interval by declaring that permits cannot be banked or borrowed after that period and by adjusting permit allocations in the last period to offset the bank.[16] Note that the regulator will be using all the information up until period T to do this – the beginning of period bank B_T as well as the beginning of period expectation of the cost shock $E_{T-1}[\theta_T]$.

In particular, the market equilibrium given by Equation (1) is replaced by

$$y_T = (\bar{e}_T - a_T) - B_T, \tag{6}$$

where the regulator constrains B_{T+1} to zero. As before a_T is aggregate abatement, $\bar{e}_T$ is aggregate emissions before abatement, and y_T is the supply chosen by the regulator. In particular, the regulator will choose y_T so that $E_{T-1}[p_T] = p_T^*$. How? Imagine for simplicity the case of a single firm. In this case, let a_T^* be defined by

$$E_{T-1}\left[\frac{\partial C(a_T^*; \theta_T)}{\partial a_T}\right] = p_T^*. \tag{7}$$

Now let the regulator choose $y_T = (\bar{e}_T - a_T^*) - B_T$. That is, the regulator chooses a permit supply level that, in conjunction with the beginning of period bank B_T and uncontrolled emissions $\bar{e}_T$, implies a level of abatement that delivers the expected price. As discussed further at the end of this section, while a_T^* is the optimal *expected* abatement in the last period given the desire to hit price p_T^* , it is not the optimal level once the cost shock is revealed. As firms are nonetheless required to achieve a_T^* in the last period, the actual market price in period T when the shock θ_T is realized, will not equal p_T^* but will be given by

$$\frac{\partial C(a_T^*, \theta_T)}{\partial a_T} = p_T. \tag{8}$$

Note that Equation (7) defines the best guess about the desired abatement level a_T^* at period $T-1$ given incomplete information about θ_T, while equation (8) indicates the resulting price level when the regulator fixes $a_T = a_T^*$ and the actual cost shock θ_T is revealed.

By choosing y_T so that the expected value $E_{T-1}[p_T] = p_T^*$, the regulator can establish the desired price, on average, in period T and, more importantly, the expected price prior to period T. Having chosen $R_t = p_t^*/(\beta^* p_{t+1})$ to satisfy Equation (5) based on the desired price path $\{p_t^*\}$, the regulator establishes the desired price, $p_t = p_t^*$, with certainty in all periods $t < T$. That is, (5) coupled with $E_{T-1}[p_T] = p_T^*$ implies $p_{T-1} = \beta R_{T-1} E_{T-1}[p_T] = p_{T-1}^*$, $p_{T-2} = \beta R_{T-2} E_{T-2}[p_{T-1}] = p_{T-2}^*$, and so on. A more formal proof is provided in Appendix A.

Note that by subtracting the period T bank from the allocation in period T, it does not really matter what the regulator does in periods $t < T$ concerning permit supply, so long as the bank is not too large.[17] For example, the regulator could wait and issue all the allowances for the entire policy horizon in the last period by allowing firms to borrow against that final allocation. In such a ‘‘true-up’’ period, any accumulated bank or debt is offset by a more generous (stingy) permit allocation for that period. In this manner, quantity shocks that have accumulated in the bank are absorbed by the regulator’s allocation, y_T. Thus, our ‘‘quantity-plus’’ policy replicates a price policy through intertemporal quantity arbitrage and banking/borrowing, and by adjusting for any cumulative abatement surplus or shortfall through a final period allocation rule that adjusts for unexpectedly low or high costs.

In effect, the regulator accepts the information asymmetry that reveals θ_t with a lag (the regulator must wait to observe the market price and abatement level). However, rather than fixing the permit level for many periods, the regulator uses newly revealed information to set the permit supply in each period (or at least in the very last period). The only ‘‘error’’ in terms of missing the desired price path, p_t^* , occurs in the final period, T, where $p_T = p_T^* + f(\theta_T)$ and $E[f(\theta_T)|\{\theta_1, ..., \theta_{T-1}\}] = 0$.[18] If autocorrelation in θ_t is high and/or the periods are short enough (e.g., quarters rather than years), this error may be very small. This approach can be applied repeatedly in the case of an infinite regulatory horizon.

These results can be related back to the literature on banking and borrowing without uncertainty. With no uncertainty, achieving the desired price path implies a restriction only on the trading ratio and total permit supply, not supply in each period. In the present model, where price uncertainty in all periods but the last is eliminated in equilibrium, the two parallel necessary restrictions to achieve the desired price path are the trading ratio and permit supply in the last period $(y_T = (\bar{e}_T - a_T^*) - B_T)$. The amount of banked permits at the beginning of the last period reflects the realized values of all prior cost shocks as well as (arbitrary) choices about permit supply in previous periods. By adjusting permit supply in the last period by the amount of banked permits the policy effectively wraps up all the prior uncertainty into the bank and compensates for it ex post.

3.2. FIXING PRICES BY ADJUSTING ALLOCATIONS BASED ON PAST ABATEMENT AND PRICES

An alternative approach to achieving a price target in the case of an infinite horizon (or an additional instrument within a finite horizon) is to specify a permit supply rule that periodically adjusts allocations based on observations of past prices and abatement to account for past cost shocks. Before laying out the rule, it is useful to first describe the regulator’s optimal permit supply rule and optimal aggregate abatement level in the absence of cost shocks. In particular,[19]

$$\frac{\partial C_i(a^*_{i,t},0)}{\partial a_{i,t}} = p^*_t \, \forall i;\; a^*_t = \sum_i a^*_{i,t} \quad \text{and} \quad y^*_t = \bar{e}_t - a^*_t. \tag{9}$$

In other words, at each point in time, each firm's marginal costs equal the desired permit price, individual abatements sum to aggregate abatement, and the supply of permits equals the chosen residual level of emissions. Note that without cost shocks, banking and borrowing would not be necessary since the regulator has enough information to use quantity controls to correctly fix the price in each period.

Having defined the optimal permit allocation in the absence of cost shocks (y^*_t), the permit allocation rule in the presence of cost uncertainty is

$$\begin{aligned} y_0 &= y^*_0, \\ y_{t+1} &= y^*_{t+1} - R_t\left[a_t(p^*_t,\theta_t) - a^*_t\right], \quad t \geq 0, \end{aligned} \tag{10}$$

where the adjustment to the permit allocation ($y_{t+1} - y^*_{t+1}$) is simply equal to the previous period's aggregate uncertainty-related quantity shock, brought forward one period by the trading ratio (R_t). Here, that quantity shock is represented as the abatement that would occur at the desired price allowing for the uncertain shock, $a_t(p^*_t,\theta_t)$, minus the abatement level at that price absent any shock, a^*_t from (9). In other words, the regulator specifies a permit supply rule that fixes prices by adjusting allocations to exactly offset realized cost shocks, based on those prices.

In practice, the regulator does not observe $a_t(p^*_t,\theta_t)$ or θ_t, but instead must infer their value by observing the actual values of a_t and p_t. To operationalize (10), we rewrite $a_t(p^*_t,\theta_t) - a^*_t$ and then make a linear approximation:

$$\begin{aligned} a(p^*_t,\theta_t) - a^*_t &= a(p^*_t,\theta_t) - a(p^*_t,0) \\ &= a(p_t,\theta_t) - a(p^*_t,0) - (a(p_t,\theta_t) - a(p^*_t,\theta_t)) \\ &= a_t - a^*_t - (p_t - p^*_t)\sum_i \left(\frac{\partial^2 C_i(a^*_{i,t},0)}{\partial a^2_{i,t}}\right)^{-1}. \end{aligned} \tag{11}$$

Here we have made use of the fact that

$$a_i(p_t,\theta_t) - a_i(p^*_t,\theta_t) = (p_t - p^*_t)\frac{\partial^2 C\,(a^*_{i,t},0)}{\partial a^2_{i,t}} \tag{12}$$

based on a linear approximation to (3), and then aggregated across firms to derive the last term in the last line of (11).

This leads to the following revised supply rule with uncertainty,

$$y_0 = y_0^*,$$
$$y_{t+1} = y_{t+1}^* - R_t \left[a_t - a_t^* - (p_t - p_t^*) \sum_i \left(\frac{\partial^2 C_i(a_{i,t}^*, 0)}{\partial a_{i,t}^2} \right)^{-1} \right], \quad t \geq 0 \tag{13}$$

Equation (13) allows for the fact that the market price p_t might not equal the desired price p_t^*. In this case the supply adjustment, which is meant to reflect the effect of the cost shock, will equal the difference between observed and desired abatement, *minus* any difference in abatement owing to an incorrect market price. If, for example, $p_t < p_t^*$, a full adjustment by $a_t - a_t^*$ would be too much – it would represent more than the increased emissions (reduced abatement) due to the cost shock. However, adjusting for $(p_t - p_t^*)\partial^2 C_i(a_{i,t}^*, 0)/\partial a_{i,t}^2$ delivers exactly the supply adjustment consistent with the price equaling the desired price.

One can view the second bracketed term in (13) as a potential penalty. Normally, when prices follow the desired path, the regulator would make up in period $t+1$ any difference between allowances used and allowances supplied in period t. That is, the regulator would decrease permit supply in period $t+1$ by the observed value of $a_t - a_t^*$(extra abatement). However, if prices do not follow the desired path, the regulator does not make up this difference based on the "penalty" term, and a borrowing debt (or bank) persists.

If a debt (or bank) develops because of this price penalty in one period, firms must hold similar expectations about future periods based on the optimization condition, $E_t[p_{t+s}] = \beta R_{t+s} E_t[p_{t+s+1}]$. Namely, if $E_t[p_{t+s}] < p_{t+s}^*$ in one period, then $E_t[p_{t+s+1}] < p_{t+s+1}^*$ and $E_t[a_{t+s+1}] < a_{t+s+1}^*$ (from (12)), and the debt will grow. Assuming a finite limit to borrowing, this limit will be reached at some point in the future.[20] At that moment, say, time $t+s$, prices will necessarily rise implying that $E_t[p_{t+s-1}] < \beta R_{t+s-1} E_t[p_{t+s}]$. Recognizing this, firms have a financial incentive to abate more now and bank permits to sell at above-market returns in period $t+s$. If firms behave this way, it will bring current prices in line with p_t^*. A more formal proof is provided in the appendix.

3.3. TIME CONSISTENCY AND COMMITMENT

Both of the above proposals require commitment by the regulatory authority. That is, the proposals define a rule for regulator behavior that either establishes $E_{T-1}[p_T] = p_T^*$ or punishes undesired price paths through limits on borrowing. When the time to implement the commitment arrives in the first case, or if firms challenge the regulator's commitment in the second case, the regulator will want to break the commitment based on current period costs and benefits. More specifically in the case of the first policy, which adjusts allocations and closes the bank in the final period, the regulator will eventually want to relax the banking/borrowing constraint that $B_{T+1} = 0$ and allow banking and borrowing in the last period. Otherwise, the actual price in the last period will deviate from the desired price by some random amount. As noted earlier, this deviation may be negligible if the regulator is able to collect information up to period T and make an accurate forecast of θ_T.

In the case of the second policy of periodic allocation adjustments, if the price, p_t, falls below the desired price, p_t^*, eventually a permit shortage appears as demand exceeds supply. In particular, borrowing will occur that eventually breaches any finite limit. At that point, say period $t + s$, the price must return to p_{t+s}^* to meet the borrowing constraint on average and exceed p_{t+s}^* if there are any adverse shocks.[21] Yet, even if firms box themselves into this corner, the regulator will not at that point want to punish them by allowing the price to be higher than the desired level; the optimal response moving forward is to forgive, allow borrowing, and try to avoid missing the price target in the next period.

As in the literature on monetary policy, the regulator cannot set rules break them, and then expect firms to fall for the same trick a second (or third) time. Continuing the analogy, there exists a choice between strictly enforcing the rules and finding a discretionary solution. In a discretionary solution the regulator balances the immediate welfare gain from some degree of "forgiveness" against the decreased effort by firms to achieve the chosen price path (in the monetary policy example, the authority is balancing an immediate output gain against persistent inflation). But the most relevant question for us is whether we can expect a regulatory authority to commit to these kinds of rules.

In reality, regulatory authorities successfully undertake such commitments all the time. U.S. family welfare payments are stopped after specified time limits, sometimes regardless of the recipient's current situation. Criminals are incarcerated for long periods under "three-strikes you're-out" policies, even if the third strike is a misdemeanor. It seems that the broader, longerterm negative effects of not carrying out a commitment carry real weight actual policy settings. Nonetheless, the fact that each of the above policies depends on future commitments rather than current action suggests it might be more fruitful to opt for real-time policy actions that could be applied on an as-needed basis.

3.4. CONTEMPORANEOUS INSTRUMENTS FOR FIXING PRICES

Each of the above approaches relies on firms properly taking into account expectations of future action, potentially leading to mistakes as well as commitment problems. In this section, we instead propose actions that take place contemporaneously with cost shocks to immediately offset their consequences and stabilize prices. Each of the methods below represents a different approach for achieving the same end – stabilizing permit prices by adjusting the permit cap in response to cost fluctuations. While we have not formalized the policies or resulting behavior, we speculate about their consequences and believe they could be fruitful areas for further analysis.

Direct government buying and selling of permits. Perhaps the most direct method to attain a particular price target would be for the regulatory authority to regularly intervene through direct buying and selling of permits at the target price, much like Federal Open Market Committee uses market transactions of government securities to influence interest rates. Knowing that regulators will intervene to maintain a particular price level, firms would abate, bank, and borrow in accordance with that expectation, thereby fixing the price path. While this is a potentially attractive option to keep in mind, it compromises our original intent to entertain only methods that do not involve monetary exchanges between the regulatory authority and firms (note that combined with the finite-horizon borrowing policies noted earlier it might be possible to limit the need for government sales to unexpected shocks in the last period).

Adjustable permit reserves. With this option, at least some – but not necessarily all – regulated firms would be subject to a reserve requirement to always hold certain quantity of unused permits in their accounts. For example, the reserve requirement could be imposed on all firms receiving gratis allocations, based on a certain percentage of the allocation. Or, the requirement could be imposed on current emitters based on a certain percentage of last year's emissions. In both cases, initial reserves could be created through a special allocation process. These reserves would be roughly analogous to the reserve requirement that the Federal Reserve places on banks, whereby they are required to always hold and not loan out certain percentage of deposits.[22] As with the Fed's reserve requirement, firms not meeting the permit reserve requirement could be allowed to borrow from the regulatory authority in order to meet it.

Although the reserves are held by the regulated firms, the regulator maintains control over the use of the reserves. This gives the regulator an additional policy lever to stabilize permit prices by influencing the effective amount of permits in circulation, in the same manner that the Fed can adjust reserve requirements to influence the interest rate. Raising the reserve requirement, for example, would lower the effective amount of permits available in the market, thereby raising the permit price. Lowering the reserve requirement would have the opposite effect. The regulator could take this action any time it saw prices deviating from the target.

As with direct government intervention, this option would encourage permit buying and selling in response to cost fluctuations, but it would effectively delegate this responsibility to permit holders through the use of permit reserves. Because firms holding reserves would know in advance that the regulator will react to prices above or below the target by adjusting the reserve requirement, these firms will have an incentive to act first. Irrespective of whether the regulator actually adjusts the reserve requirement, they will want to acquire permits when the price falls below the target in anticipation of the government requiring them to do so, potentially at a less favorable price (and the reverse if costs are high).

In effect, to avoid monetary exchanges with the regulated firms, the regulator simply places the pool of "potential permits" in the hands of the firms, and the firms themselves do all the buying and selling – under the watchful eye (and perpetual threat) of the intervening regulator. Borrowing and banking would be equivalent to holding permits below or above the reserve requirement. Here, the level of the reserve requirement acts as an effective limit on borrowing, a requirement noted earlier for the supply rule (13).

Loans and special allocations. All of the preceding approaches allow for some form of regular, borrowing above the expected cap level if costs are unexpectedly high – a policy that could be unwelcome to environmentalists who want emissions limits treated as rigid constraints. An alternative would be to allow the regulator to react to specific high-permit-price circumstances by making special allocations. That is, when the permit price reaches a particular threshold, the regulator could give away some volume of additional permits, thereby lowering permit prices. To be equitable and avoid a situation where some permits are given away shortly after others have been sold at a high price, it might be preferable to loan rather give away these permits. These loans could be distributed through a bidding process whereby firms bid the interest rate they would be willing to pay on these special loans, thereby providing a simple and fair distribution mechanism for the extra permits.

Splits and reverse splits. A direct means of influencing prices is to directly change the quantity of outstanding permits by announcing a split or reverse split of existing permits, as with shares in the stock market. If prices are too high, for example, the regulator can simply announce an X-for-1 split so that one unit of permits is converted into X units. A reverse split of 1-for-X permits could be used if permit prices were too low.

4. Concluding Remarks

Many competing forces determine the design of environmental policy. Often, these forces lead to quantity-based regulation despite a large expected gain from price-based controls (Newell and Pizer 2003). In the case of efforts to mitigate the consequences of global climate change, this tendency toward quantity-based regulations leads to concern over the uncertain costs of particular targets, with banking and especially borrowing arising as a potential solution.[23]

We demonstrate that permit systems incorporating banking, borrowing, and adjustments to the quantity of outstanding permits can replicate price-based regulation. The methods do not require any monetary transfers between the government and the regulated firms, thereby avoiding a politically unattractive aspect of price-based policies. The approaches we lay out can work for both finite and infinite horizons and involve a variety of instruments ranging from adjustment of allocations based on past prices and abatement to the establishment of adjustable permit "reserves," splits and reverse splits, loans, and special allocations. With such a wide range of potential options, opposition to overt price policies should not be viewed as an obstacle to considering other means of achieving the flexibility associated with these instruments.

Acknowledgements

Newell and Pizer are Fellows at Resources for the Future, Washington, DC. Zhang is an Economist at the Asian Development Bank, Manila. We thank Matti Liski and two anonymous referees for helpful comments on a previous version. The opinions expressed in this paper do not necessarily reflect the views of the Asian Development Bank.

Notes

1. Examples of US permit programs that have allowed banking include ones to curb criteria pollutants under the Clean Air Act, phase down lead in the 1980s trade sulfur dioxide emissions under Title IV of the Clean Air Act Amendments of 1990 (Schennach 2000), trade nitrogen oxide emissions in the northeast states, and reduce NO_X and particulate emissions from heavy-duty truck and bus engines.
2. Corporate Average Fuel Economy standards allow car manufacturers to both bank and borrow fuel economy credits for up to three years. California's Low-Emission Vehicle Program also allows vehicle manufacturers to receive debits to be made up in the following model year (Rubin and Kling 1993). International climate policy discussions have implicitly included borrowing within possible consequences for noncompliance under the Kyoto Protocol, through the payback of excess tons with a penalty (i.e., interest) (United Nations Framework Convention on Climate Change 2000).
3. At the international level, permits may avoid problems associated with international tax payments (Wiener 1998). Furthermore, if intertemporal trading supplants intratemporal trading as the main source of compliance flexibility among countries, this would weaken concerns about international capital flows and balance of payments caused by international permit trades (McKibbin et al. 1999).

4. Note this is true even with uncertainty about benefits when the benefit uncertainty is not revealed before emissions are determined. In that case, it is still first-best from the vantage point of achievable policy outcomes.
5. Also see Requate (2002) and Phaneuf and Requate (2002) on the influence of cost uncertainty on the level and welfare effects of permit banking.
6. Yates and Cronshaw (2001) also solve for the optimal intertemporal trading ratio in a two-period model, finding that, in addition to the discount rate, the ideal trading ratio is a function of the parameters of the cost and benefit functions, including the degree of cost uncertainty.
7. The direct buying and selling of permits at fixed prices by the government would be the most straightforward way to implement an arbitrary price policy – a point we revisit a bit later.
8. The allocation need not be for the immediately subsequent period; for example, business interests typically advocate allocation in blocks of 5–10 years. Our results are easily extended in this case, although it might suggest additional interest in the contemporaneous interventions noted in Section 3.4.
9. While it might seem like the trading ratio would be a useful policy lever to manage permit prices, it is not. As we will see shortly, the trading ratio is what anchors the relations among future permit prices and, given our desire to keep future permit prices fixed, must also remain fixed.
10. Here and throughout, the absence of an i subscript indicates aggregate variables; e.g., a_t is aggregate abatement and $a_{i,t}$ is abatement by firm i (both at time t).
11. The assumption that the cost shock is the same for all firms can be relaxed without changing the basic results.
12. We have not included the cost of permits for uncontrolled emissions, $p_t\bar{e}_{i,t}$, because it does not affect the choice of abatement or banking.
13. That is, current period shocks affect current prices only to the extent that they shift the entire expected price path. This might be true, for example, in the case of a permanent shift in costs.
14. Note that we have not explicitly optimized over environmental damages. Rather, we assume that the social welfare function leads to a preference for a particular price path, and we simply try to achieve that price path. For economic optimality, prices would correspond to marginal damages.
15. A subtle question is whether we fix $E_t[p_T] = p_T^*$ or $E_{T-1}[p_T] = p_T^*$, the difference being new information that might arrive between time t and $T-1$. Because p_T and the regulator's rules are both fixed, we do not anticipate any change in relevant information and these two expectations ought to be equal.
16. The permit system may end in period T, banking/borrowing may simply cease, or the policy may continue with a new system of intertemporally tradable permits, lasting from $T+1$ to $2T$, for example, that are not exchangeable with the earlier system.
17. That is, the bank cannot be so large that $y_T < 0$; for credibility, it may be desirable to keep the bank relatively small.
18. Thus the final period looks like the traditional one-period efficiency divergence between prices and quantities.
19. This establishes the relationship between desired price, p^*, and quantity, a^*, absent a cost shock. The quantity a^* will equal the expected quantity if marginal cost is linear in a and θ.
20. Note that with no constraints on borrowing, firms could choose to abate nothing and borrow permits each period to cover both previously borrowed permits and new emissions. This is consistent with $p_t = 0$ for all t and satisfies (5). This behavior is possible regardless of the permit supply rule. The permit supply rule therefore needs to be coupled with limits on borrowing, with those limits set large enough to accommodate potential persistent cost shocks.

21. When borrowing hits its limit in future period $t+s$, $E_t[p_{t+s}] = p^*_{t+s}$, that is, the expected price will equal the desired price. However, with no capacity to borrow, the actual price will exceed the desired price as soon as an adverse shock occurs.
22. Another analogy to Federal Reserve policy tools would be to influence permit prices by adjusting the intertemporal trading ratio, just as the Fed influences interest rates by changing the discount it charges banks for short-term loans. Increasing (decreasing) the trading ratio applied to borrowed or banked permit would tend to increase (decrease) current permit prices because the value of permits in the present is raised (lowered) relative to the future. However, this approach fails to support a desired price path $\{p^*_t\}$ because there is only one trading ratio that reflects $R_t = p^*_t/(p^*_{t+1}\beta)$.
23. See description of borrowing implicit within possible compliance outcomes in paragraph II.XV, United Nations Framework Convention on Climate Change. United Nations Framework Convention on Climate Change (2000) Procedures and Mechanisms Relating to Compliance under the Kyoto Protocol: Note by the co-Chairmen of the Joint Working Group on Compliance, Bonn, UNFCCC.
24. If the slope of marginal costs is not exactly constant, the rule will hold in approximation.
25. Note that the prices being bound are current period prices. That is, we do not want the value of allowances to vanish or explode relative to other consumer goods.
26. The argument can be reversed for the case where the banking, rather than borrowing, limit is reached.
27. The cost function is assumed to be convex, so $k_i > 0$ in (16). Also, the bank in period T is determined by the abatement in period $T-1$.

References

Barro, R. J. and D. B. Gordon (1983), 'A positive Theory of Monetary Policy in a Natural Rate Model', *Journal of Political Economy* **91**(4), 589–610.

Cronshaw, M. B. and J. B. Kruse (1996), 'Regulated Firms in Pollution Permit Markets with Banking', *Journal of Regulatory Economics* **9**, 179–189.

Haubrich, J. G. and J. A. Ritter (2000), 'Dynamic Commitment and Incomplete Policy Rules', *Journal of Money, Credit, and Banking* **32**(4), 766–784.

Hoel, M. and L. Karp (2002), 'Taxes *versus* Quotas for a Stock Pollutant'. *Resource and Energy Economics* **24**, 367–384.

Keohane, N. O., R. L. Revesz, et al. (1998), 'The Choice of Regulatory Instruments in Environmental Policy', *Harvard Environmental Law Review* **22**, 313–367.

Kling, C. and J. Rubin (1997), 'Bankable Permits for the Control of Environmental, Pollution', *Journal of Public Economics* **64**(1), 101–115.

Kydland, F. E. and E. C. Prescott (1977), 'Rules rather than discretion: The Inconsistency of Optimal Plans', *Journal of Political Economy* **85**(3), 473–491.

Leiby, P. and J. Rubin (2001), 'Intertemporal Permit Trading for the Control of Greenhouse Gas Emissions', *Environmental and Resource Economics* **19**, 229–256.

McKibbin, W. J., M. Ross, et al. (1999), 'Emissions Trading, Capital Flows and the Kyoto Protocol', *Energy Journal,* special issue.

Newell, R. and W. Pizer (2003), 'Regulating Stock Externalities Under Uncertainty', *Journal of Environmental Economics and Management* **45**, 416–432.

Phaneuf, D. J. and T. Requate (2002), 'Incentives for Investment in Advanced Pollution Abatement Technology in Emission Permit Markets with Banking', *Environmental and Resource Economics* **22**, 369–390.

Requate, T. (2002), Does the Environment Benefit from Banking Programs for Tradable Permits? Manuscript, University of Kiel, Kiel, Germany.

Roberts, M. J. and M. Spence (1976), 'Effluent Charges and Licenses Under Uncertainty', *Journal of Public Economics* **5**(3–4), 193–208.
Rubin, J. and C. Kling (1993), 'An Emission Saved is an Emission Earned: An Empirical Study of Emission Banking for Light-Duty Vehicle Manufacturers', *Journal of Environmental Economics and Management* **25**, 257–274.
Rubin, J. D. (1996), 'A Model of Intertemporal Emision Trading, Banking, and Borrowing', *Journal of Environmental Economics and Management* **31**, 269–286.
Schennach, S. M. (2000) 'The Economics of Pollution Permit Banking in the Context of Title IV of the 1990 Clean Air Act Amendments', *Journal of Environmental Economics and Management* **40**(3), 189–210.
United Nations Framework Convention on Climate Change (2000), Procedures and Mechanisms Relating to Compliance under the Kyoto Protocol: Note by the co-Chairmen of the Joint Working Group on Compliance. Bonn, UNFCCC.
Weitzman, M. L. (1974), 'Prices vs. Quantities', *Review of Economic Studies* **41(**4), 477–491.
Weitzman, M. L. (1978), 'Optimal Rewards for Economic Regulation', *American Economic Review* **68**(4), 683–691.
Wiener, J. (1998), 'Global Environmental Regulation: Instrument Choice in Legal Context', *Yale Law Journal* **677**, 108.
Williams, R. C., III (2001), Price vs. Quantities vs. Tradable Quantities. Manuscript, University of Texas, Austin, TX.
Yates, A. J. and M. B. Cronshaw (2001), 'Pollution Permit Markets with Intertemporal trading and Assymetric Information', *Journal of Environmental Economics and Management* **42**, 104–118.
Yohe, G. W. (1978), 'Towards a General Comparison of Price Controls and Quantity Controls under Uncertainty', *The Review of Economic Studies* **45**(2), 229–38.

Appendix A

PROOF FOR SECTION 3.1

In the last period, given B_T and $E_{T-1}[\theta_T]$ choose y_T such that $y_T = \bar{e}_T - a_T - B_T$ and $a_T = \sum_i f_i(p_T^*)$ where $f_i^{-1}(a) = E_{T-1}[\partial C_i(a, \theta_T)/\partial a]$. Assuming monotonic marginal costs, the unique solution to Equations (3) and (6) implies $E_{T-1}[p_T] = p_T^*$. Now choose $R_t = p_t^*/(\beta p_{t+1}^*)$. Applying this relation and Equation (5) iteratively, we have $p_t = p_t^*$.
Note that the preceding proof does not make any assumptions about y_t for $t < T$. Therefore, it does not matter how the regulator allocates allowances in periods $t < T$. (If we require that $y_T > 0$, the regulator will need to make sure that B_T is not larger than $\bar{e}_T - a_T$, where a_T is as given above.)

Proof for Section 3.2

From (5), we know that $p_t = \beta R_t E_t[p_{t+1}] = \beta^s \left(\prod_{r=0}^{s-1} R_{t+r}\right) E_t[p_{t+s}]$ for all $s > 1$. That is, in equilibrium a risk-neutral, optimizing firm must have this particular expectation about the path of prices – otherwise arbitrage opportunities exist given the trading ratios R_{t+s}. Given there is one market price today, all firms in equilibrium must have those same expectations. Now suppose the regulator commits to the allocation rule (13) and there is a limit to banking and borrowing. We first show that any price expectation except $p_t = p_t^*$ will violate the banking /borrowing constraint (with probability approaching one). We then show that such a price path that reaches the limit with certainty cannot be optimal.

Consider a price path $p_t^\dagger \neq p_t^*$ for $t \geq 0$ that satisfies the first-order conditions (Equation (5)) in each period: $p_t^\dagger = \beta R_t p_{t+1}^\dagger$. Define $a_{i,t}^\dagger$, and $a_t^\dagger$, as in (9), that is, the abatement path absent stochastic shocks, but defined by $p_t^\dagger$ rather than p_t^*. Now consider the expected bank T periods in the future. Taking (1) and summing over periods 0 to $T-1$ with an appropriate cumulative trading ratio $\prod_{s=t}^{T-1} R_s$ (to bring values forward to period T), we have

$$\sum_{t=0}^{T-1}\left(\prod_{s=t}^{T-1} R_s\right)[y_t - (\bar{e}_t - a_t)] = \sum_{t=0}^{T-1}\left(\prod_{s=t}^{T-1} R_s\right)(-B_t + (B_{t+1}/R_t)) = -B_0\left(\prod_{s=0}^{T-1} R_s\right) + B_T. \tag{A.1}$$

In words, Equation (A.1) simply accumulates the difference between permits supply y_t and demand $\bar{e}_t - a_t$ We can replace the permit supply y_t in (A.1) with the proposed supply rule in 13 and the definition in (9), $y_t^* = \bar{e}_t - a_t^*$. Rearranging yields

$$\sum_{t=0}^{T-1}\left(\prod_{s=t}^{T-1} R_s\right)[a_t - a_t^*] + \sum_{t=1}^{T-1}\left(\prod_{s=t-1}^{T-1} R_s\right) \times \left[-(a_{t-1} - a_{t-1}^*) + (p_{t-1} - p_{t-1}^*)\sum_i \left(\frac{\partial^2 C_i(a_{i,t-1}^*, 0)}{\partial a_{i,t-1}^2}\right)^{-1}\right] = B_0\left(\prod_{s=0}^{T-1} R_s\right) + B_T. \tag{A.2}$$

We simplify (15) further by first assuming $\frac{\partial^2 C_i(a,0)}{\partial a^2} = k_i$ (i.e., treating it as constant)[24] and recognizing that $p_{t+s-1} = p_{t+s-1}^\dagger$ under the maintained hypothesis we are attempting to disprove. From (9), we then have

$$(a_t^\dagger - a_t^*) = (p_t^\dagger - p_t^*)\sum_i k_i^{-1}. \tag{A.3}$$

That is, absent shocks the deviation in abatement by firm i is related to the deviation in price by k_i^{-1}, and these deviations can be summed to yield the aggregate deviation. Equation (A.2) can now be rewritten as

$$\sum_{t=0}^{T-1}\left(\prod_{s=t}^{T-1} R_s\right)[a_t - a_t^*] - \sum_{t=1}^{T-1}\left(\prod_{s=t-1}^{T-1} R_s\right)[a_{t-1} - a_{t-1}^\dagger] = -B_0\left(\prod_{s=0}^{T-1} R_s\right) + B_T. \tag{A.4}$$

In words, the change in the bank from time 0 to time T is comprised of two components, the sum of the observed deviations from the desired abatement path absent stochastic shocks, $a_t - a_t^*$, *minus* a supply correction arising from observed stochastic shocks. The correction, which lags one period, is based on the deviation from the expected abatement leve $a_t^\dagger$ absent stochastic shocks, given the observed prices $p_t^\dagger$ under the maintained hypothesis. This correction arises because the proposed permit supply rule (13) first allocates based on the desired abatement path, $y_t^* = \bar{e}_t - a_t^*$, and then *corrects* each period for changes in emissions arising from previous period cost shocks. We can simplify (A.4) further because all of the deviations from $a_t^\dagger$ except the last one are corrected:

$$R_{T-1}(a_{T-1} - a_{T-1}^*) + \sum_{t=0}^{T-1}\left(\prod_{s=t}^{T-1} R_s\right)(a_t^\dagger - a_t^*) = B_T, \tag{A.5}$$

where here we have, without loss of generality, also made the simplifying assumption that the bank starts at zero, $B_0 = 0$.

Equation (A.5) is crucial in this proof because it establishes that with the proposed supply rule (13), stochastic shocks wash out and the *actual* future bank depends entirely on the *expected* future bank, except for the last period shock embodied in the first term.

Based on this relation, we can see that whether or not the bank exceeds an arbitrary borrowing or banking constraint depends on the following limit (below we see that the first term in (A.5) introduces a probabilistic convergence but otherwise can be safely ignored):

$$\lim_{T\to\infty}\sum_{t=0}^{T-1}\left(\prod_{s=t}^{T-1} R_s\right)(a_t^\dagger - a_t^*). \tag{A.6}$$

This expression represents the behavior in the limit of the sum of the expected abatement deviation, based on the proposed price path. Using (A.3) we can write the object of the limit in (A.6) in terms of price deviations rather than abatement deviations:

$$\sum_{t=0}^{T-1}\left(\prod_{s=t}^{T-1} R_s\right)(a_t^\dagger - a_t^*) = \sum_{t=0}^{T-1}\left(\prod_{s=t}^{T-1} R_s\right)(p_t^\dagger - p_t^*)\sum_i k_i^{-1}. \tag{A.7}$$

Further, note that the arbitrage equation (5) applies to both the optimal price path p_t^* and hypothesized path $p_t^\dagger$, and therefore their difference. Applied iteratively it yields

$$(p_0^\dagger - p_0^*) = \beta^t\left(\prod_{s=0}^{t-1} R_s\right)(p_t^\dagger - p_t^*), \tag{A.8}$$

allowing us to rewrite the expression (A.7) as

$$\sum_{t=0}^{T-1}\left(\prod_{s=t}^{T-1} R_s\right)(a_t^\dagger - a_t^*) = \sum_{t=0}^{T-1}\left(\prod_{s=t}^{T-1} R_s\right)\left(\prod_{s=0}^{t-1} R_s\right)^{-1}\beta^{-t}(p_0^\dagger - p_0^*)\sum_i k_i^{-1}. \tag{A.9}$$

We now require an assumption that the desired price path does not tend to zero or infinity – that is,

$$\text{there exist } \underline{p}, \bar{p} > 0 \quad \text{such that } \bar{p} \geq P_t^* \geq \underline{p} \text{ for all } t. \tag{A.10}$$

This assumption allows us to bound the terms involving R_s and β in (A.9).[25] Consider now the implication of the arbitrage equation (5) coupled with (A.10). Namely,

$$\frac{\bar{p}}{p_0^*} \geq \beta^t\prod_{s=0}^{t-1} R_s \geq \frac{\underline{p}}{p_0^*}. \tag{A.11}$$

Given $0 < \beta \leq 1$ and (A.11),

$$\left(\prod_{s=t}^{T-1} R_s\right)\left(\prod_{s=o}^{t-1} R_s\right)^{-1}\beta^{-t} \geq \left(\prod_{s=t}^{T-1} R_s\right)\beta^{T-t}\left(\prod_{s=o}^{t-1} R_s\right)^{-1}\beta^{-t}$$
$$\geq \left(\prod_{s=0}^{T-1} R_s\right)\beta^{T}\left[\left(\prod_{s=o}^{t-1} R_s\right)^{-1}\beta^{-t}\right]^2$$
$$\geq \frac{\underline{p}}{p_0^*}\left(\frac{\bar{p}}{p_0^*}\right)^{-2}$$
$$\geq \frac{\underline{p}\cdot p_0^*}{\bar{p}^2}. \tag{A.12}$$

This allows us to establish the magnitude of the expression in (A.9),

$$\left|\sum_{t=0}^{T-1}\left(\prod_{s=t}^{T-1} R_s\right)(a_t^\dagger - a_t^*)\right| \geq \sum_{t=0}^{T-1}\frac{\underline{p}\cdot p_0^*}{\bar{p}^2}\left|(p_0^\dagger - p_0^*)\sum_i k_i^{-1}\right|$$
$$\geq T\frac{\underline{p}\cdot p_o^*}{\bar{p}^2}\left|(p_0^\dagger - p_0^*)\sum_i k_i^{-1}\right|, \tag{A.13}$$

where the expression multiplying T is a lower limit on the magnitude of the contribution to a future bank from the expected abatement deviation in each period based on the proposed alternate price path.

We now return to the first term in (A.5), which we had previously set aside. Taking a full differential around the optimum from (9) we have

$$p_t^\dagger - p_t^* = \frac{\partial^2 C_i}{\partial a^2}(a_{i,t} - a_{i,t}^*) + \frac{\partial^2 C_i}{\partial a \partial \theta}\theta_t,$$
$$a_t - a_t^* = \left(\sum_i k_i^{-1}\right)(p_t^\dagger - p_t^*) - \left(\sum_i k_i^{-1}\frac{\partial^2 C_i}{\partial a \partial \theta}\right)\theta_t, \tag{A.14}$$

where second line solves for the differential change in abatement, then sums over firms. From (A.8) and (A.11) we know that $p_t^\dagger - p_t^*$ is bounded by

$$\left|p_t^\dagger - p_t^*\right| \leq \left|p_0^\dagger - p_0^*\right|\frac{p_0^*}{\underline{p}}.$$

With the assumption that θ_t is stationary, we know there exists an M_ε for every ε such that $Pr[|\theta_t| > M_\varepsilon] < \varepsilon$. Given that

$$\left|\left(\sum_i k_i^{-1}\right)(p_t^\dagger - p_t^*) - \left(\sum_i k_i^{-1}\frac{\partial^2 C_i}{\partial a \partial \theta}\right)\theta_t\right|$$
$$\leq \left|\left(\sum_i k_i^{-1}\right)(p_t^\dagger - p_t^*)\right| + \left|\sum_i k_i^{-1}\frac{\partial^2 C_i}{\partial a \partial \theta}\right||\theta_t|$$
$$\leq \left|\sum_i k_i^{-1}\right|\left|p_0^\dagger - p_0^*\right|\frac{p_0^*}{\underline{p}} + \left|\sum_i k_i^{-1}\frac{\partial^2 C_i}{\partial a \partial \theta}\right||\theta_t|,$$

we have

$$Pr\left[\left|\left(\sum_i k_i^{-1}\right)(p_t^\dagger - p_t^*) - \left(\sum_i k_i^{-1}\frac{\partial^2 C_i}{\partial a \partial \theta}\right)\theta_t\right| < \left|\sum_i k_i^{-1}\right|\left|p_0^\dagger - p_0^*\right|\frac{p_0^*}{\underline{p}} + \left|\sum_i k_i^{-1}\frac{\partial^2 C_i}{\partial a \partial \theta}\right| \cdot M_\varepsilon\right] > 1 - \varepsilon. \quad (A.15)$$

In other words, with any level of certainty we desire $(1-\varepsilon)$, we can establish a maximum magnitude for the first term in (A.5) based on (A.15). Given the second term is rising deterministically and linearly in magnitude, it will eventually dominate the first term. Therefore, any fixed banking/borrowing limit will be breached with probability one unless the second term vanishes (which only occurs if $p_t^\dagger = p_t^*$ for all t). Expressed formally, for any level of certainty $(1-\varepsilon)$, maximum banking/borrowing limit $B_{\max}$, and $p_0^\dagger \neq p_0^*$ define M_ε as above and choose $T < \infty$ such that

$$T\frac{\underline{p}\cdot p_0^*}{\bar{p}^2}\left|(p_0^\dagger - p_0^*)\sum_i k_i^{-1}\right| - R_{T-1}\left(\left|\sum_i k_i^{-1}\right|\left|p_0^\dagger - p_0^*\right|\frac{p_0^*}{\underline{p}} + \left|\sum_i k_i^{-1}\frac{\partial^2 C_i}{\partial a \partial \theta}\right| M_\varepsilon\right) > B_{max}.$$

From (A.5)

$$|B_T| \geq \left|\sum_{t=0}^{T-1}\left(\prod_{s=t}^{T-1} R_s\right)(a_t^\dagger - a_t^*)\right| - \left|R_{T-1}(a_{T-1} - a_{T-1}^*)\right| \quad .$$

And from (A.13) and (A.15)

$$Pr\left[\begin{array}{c}\left|\sum_{t=0}^{T-1}\left(\prod_{s=t}^{T-1} R_s\right)(a_t^\dagger - a_t^*)\right| - \left|R_{T-1}(a_{T-1} - a_{T-1}^*)\right| > \\ T\,\frac{\underline{p}\cdot p_0^*}{\bar{p}^2}\left|(p_0^\dagger - p_0^*)\sum_i k_i^{-1}\right| - \left(\left|\sum_i k_i^{-1}\right|\left|p_0^\dagger - p_0^*\right|\frac{p_0^*}{\underline{p}} + \left|\sum_i k_i^{-1}\frac{\partial^2 C_i}{\partial a \partial \theta}\right| M_\varepsilon\right)\end{array}\right] > 1 - \varepsilon.$$

Therefore $\Pr[|B_T| > B_{max}] > 1 - \varepsilon$.

To see that reaching the banking/borrowing limit with certainty is not optimal, consider the decision to borrow each period when, at some point in the future time T, borrowing will become constrained because the debt in period $T+1$ exceeds the limit.[26] The accumulation of a large borrowing debt implies that abatement has been less than desired. That is, from (A.5) the borrowing limit $(B_T \ll 0)$ is reached when $a_t^\dagger - a_t^* < 0$. From (16), this implies that $p_t^\dagger - p_t^* < 0$ until the borrowing limit is reached, e.g., time $t < T$.[27] Once the borrowing limit is reached at time, e.g., time $t \geq T$. , abatement *must* equal (or exceed) uncontrolled emissions $\bar{e}_t$, minus the allocation y_t , because borrowing is no longer an option. Given the expected allocation in period t is y_t^*, the expected abatement at this point will be the desired abatement $a_t^* = \bar{e}_t - y_t^*$ and the expected price will be the desired price p_t^*. Then we have $p_{T-1} = p_{T-1}^\dagger < p_{T-1}^* = \beta R_{T-1} p_T^* = \beta R_{T-1} E[p_T]$.

But if $p_{T-1} < \beta R_{T-1} E[p_T]$, can borrowing be desirable in period $T-1$ at which point it remains unconstrained (e.g., choosing the bank B_T)? This contradicts the first-order condition (4) (and specifically (5)) for the optimum when banking is unconstrained. Therefore, such a situation cannot be optimal.

In words, it cannot be optimal to borrow when prices are unusually low *if* we expect prices to rise. In that case, *banking* is desirable to take advantage of the arbitrage opportunities associated with rising prices. Here we have shown that

(a) prices below the desired level coupled with the chosen supply rule (13) eventually lead to an arbitrarily large borrowing debt;
(b) if there is a limit to borrowing 1 banking, this limit will be triggered by such prices at which point prices will necessarily rise, in expectation, to the desired level (or higher);
(c) this cannot be an optimal response.

If firms see unexpectedly low allowances prices, the optimal response must be to bank allowances, bidding the allowance price up to the desired (and only perfect-foresight equilibrium) level and removing arbitrage opportunities.