DISCUSSION DRAFT: PLEASE DO NOT QUOTE WITHOUT PERMISSION

*Banking and price containment in California's greenhouse gas emissions market: an experimental analysis of market design**

William Shobe; Charles Holt
University of Virginia

Thaddeus Huetteman
PEAR, Inc.

Version date: July 31, 2013

* The authors wish to acknowledge the guidance of Michael Gibbs and his colleagues at the California Air Resources Board. We received helpful comments from Kedin Kilgore, Andy Van Horn, John Melby, Jan Mazurek, Jan Grygier, and Dallas Burtraw. Key financial support for this research was provided by the U.S. Environmental Protection Agency, by a consortium of California funders lead by Pacific Gas & Electric Company, and by the U.Va. Bankard Fund. Thanks also to Sijia Yang for her tireless and enthusiastic research assistance.

Abstract

We use a set of economic experiments to test the effects of two key features of California's new program for limiting greenhouse gas emissions. The cap & trade scheme included in the program includes two novel features, limits on allowance ownership (or 'holding limits') and a tiered price containment reserve sale. These program features are linked by their potential to affect liquidity in the market for emission allowances. We examine the effects of these features on liquidity and on measures of market performance including efficiency, price discovery, and price variability. We find that tight holding limits have the effect of substantially lowering the number of banked allowances available for trade, hence lowering liquidity. This impairs the ability of traders to smooth prices over time resulting in lower efficiency, less effective price discovery, and higher variability in price. The price containment reserve, while increasing the supply of allowances available to traders, does not appear to mitigate the effects of tight holding limits on market outcomes. As a result, the imposition of holding limits in the allowance market may have the consequence of increasing the likelihood of the market manipulation that they were intended to prevent.

1. Introduction

1.1. California GHG emission regulations under AB 32

According to recent data for 2010 from the U.S. Energy Information Administration (U.S. Energy Information Administration, 2013), the state of California ranked 22nd in CO_2 emissions among the world's political jurisdictions. That placed it somewhere between Spain and France. This was true even after its emissions had fallen by more than 10 per cent from those of a decade earlier. So it is no small matter to global greenhouse gas (GHG) emissions that the state has begun implementing what arguably will, when fully implemented, be the most comprehensive set of GHG controls anywhere on the planet. The program, which is often referred to as "AB 32" after its 2006 authorizing legislation: the Global Warming Solutions Act (California, 2006), will result in 2020 emissions in California reduced to 1990 levels and 2050 emissions 20 percent below that. That California emissions per person are already the lowest in the U.S. reinforces the aggressiveness of this program.

California has implemented its GHG emissions limits with a hybrid approach that uses both regulations and a "cap and trade" scheme. Of the anticipated 80 million tons (CO_2 equivalent) per year below the business as usual forecast, the California Air Resources Board (CARB) forecasts that roughly 60 million tons of reduction will come from regulations and other programs (CARB refers to these as "complementary measures"),[1] meaning that, without the cap, the 2020 goal would be exceeded by about 20 million tons per year.

The cap, once fully implemented in 2015, will cover "up-stream" sources including power plants, large industrial facilities, and fuel suppliers, comprising some 350 separate businesses operating more than 600 separate facilities. The cap will cover approximately 85 percent of GHG emissions from the state, including those emissions imputed to imported sources of energy.

In this paper, we use a set of laboratory experiments to test the likely effect of two key design elements of the California program, limits on allowance ownership and price containment reserve sales. For these experiments, we use financially motivated human subjects who have had previous experience in a similar experimental setup.

1.2. The allowance market

The design of the California cap and trade program uses many features used in earlier schemes such as the European Union Emission Trading Scheme (EUETS), the Regional Greenhouse Gas Initiative (RGGI), and the U.S. SO2 allowance trading program under Title IV of the Clean Air Act. The cap declines over time, 2 percent per year for two years and 3 percent after that. Emission allowances representing one ton of carbon dioxide equivalent emissions (CO_2e) must be retired for each ton of emissions. Allowances are uniform except that most allowances have a vintage year, which specifies the first emission year for which they can be used. A small share of allowances will not have a vintage, but these are not tradable.

The allocation of allowances to emitters reflects a complex set of economic and political considerations. The industrial sector initially receives 90 percent of their expected need for free based on the average level of emissions per unit of output in their sector, a design chosen with an

[1] The lion's share of reductions come from a renewable electricity standard (11.5%), a low carbon fuel standard (15%), energy efficiency initiatives (12%), reducing high warming potential gases (6.5%), and forestry practices (5%).

eye to reducing incentives for production to migrate out of the state. The free share falls modestly over time. Publicly owned electric utilities receive their allowances free but must consign for auction any allowances that they do not place in their non-tradable, 'compliance account'. Investor owned utilities receive their allowances for free but must consign them to the auction and must then purchase any allowances they need, either at auction or from some other seller. CARB rules require that the publicly owned utilities use their free allocation to the benefit of ratepayers. The state will hold quarterly auctions for any allowances not allocated for free to emitters.

Covered sources must purchase each year at least 30 percent of the allowances needed to cover their emissions.[2] At the end of each three-year "compliance period", sources must retire sufficient allowances to cover all of their emissions over the three-year period. Sources in deficit after the true-up deadline must retire four allowances for each ton of emissions not covered by allowances by the compliance deadline.

CARB will hold quarterly, sealed-bid, uniform price auctions to sell any allowances not allocated for free and any allowances consigned for sale at auction. A second auction will be held on the same day to sell allowances from the vintage year three years in the future. The auctions are open to all. For the first two years covered sources are subject to the restriction that no single party may purchase more than 15 percent of the current year vintage allowances at auction and no more than 25 percent of future vintage allowances, while non-covered purchasers are limited to no more than four percent. Bids are subject to a binding reserve price starting at $10 and rising each year at 5 per cent plus the rate of inflation. Allowances not sold due to the reserve not being met are held for sale at auction after two auctions in a row have closed above the reserve price.

As of this writing, three auctions for the current, 2013, vintage of allowances have been held. The first auction, in November of 2012, closed at $10.09 per ton, just above the reserve, with the quantity bid/quantity offered ratio of 1.06. The second auction, in February of 2013 closed at $13.62 per ton with a bid to offer ratio of 2.47. The May 2013 auction closed at $14.00 with a bid to offer ratio of 1.78. The auctions for future vintages were undersubscribed and closed at the reserve price.[3]

Once the program is fully implemented, there will be a price containment reserve (PCR) amounting to approximately seven percent of the initial cap. These reserve allowances do not have a vintage, so they may be used to cover emissions from any year. CARB has chosen a novel design for releasing allowances from the price containment reserve. Six weeks after each quarterly auction, allowances will be made available in a three-tiered sale. Buyers may post offers to buy allowances at one of the three prices, starting in 2012 at $40, $45, and $50 respectively. Offers at a given price tier will be accepted at that price unless a) the next lower tier is not exhausted, in which case the buyers order will be filled at the lower tier, "pulled-down", with "pull-downs" in excess of available allowances randomly allocated, or b) if the current tier is over-subscribed, then bids in this tier will be allocated in proportion to the total number of bids at that tier. Bodsky, Donato, James, and Porter (2012) provided early experimental evidence on

[2] This feature recognizes that the possibility of bankruptcy may induce firms to under-accumulate allowances. Indeed, the only violations observed in the first compliance period of the RGGI program was from firms not accumulating allowances in anticipation of bankruptcy.

[3] Information about auction results may be found on the CARB website: http://www.arb.ca.gov/cc/capandtrade/auction/auction.htm

how these complicated rules affect bidder incentives. Later in this paper, we will investigate how this California three tier PCR sale may change incentives and outcomes as compared to the in-auction sale recently chosen by RGGI.

Cap and trade programs have, from early days, been subject to the criticism that the vertical supply curve induced by the regulatory cap may induce troublesome price variability in the allowance market, raising compliance costs. This becomes less of a problem in a multi-period market where regulated sources, anticipating price variability, can build up banks of unused allowances to use in future periods when there is a chance that prices will be much higher.[4] Banking, then, may induce responsiveness in supply, where a strictly fixed supply is offered to the market by the regulator in each period. Considerable recent research has focused on how to control excess price variability in cap and trade regimes without unduly harming the valuable price discovery properties of the allowance market (Fell, Burtraw, Morgenstern, & Palmer, 2012; Hasegawa & Salant, 2012; Stocking, 2012).

The AB 32 regulations permit the banking of unused allowances to later periods, but, in a significant break from earlier practice, CARB placed a limit on the number of allowances that a single entity could control. As we will discuss subsequently, the decision to impose limits on the ownership of allowances, referred to as holding limits, was an attempt by CARB to reduce the probability of manipulation of the allowance market.[5] The limits amount to approximately 6 million tons for 2013 and 2014, that is in the years before transportation fuels come under the cap, and 11 million tons thereafter. For a sense of the relative magnitude of this limit, Pacific Gas & Electric emissions of CO_2 for 2011 amounted to just over 2 million tons. Over a three year compliance horizon, that could put their allowable ownership under their compliance obligation. On the other hand, a relatively small emitter would not be likely to ever approach the limit.

Covered emitters can relax this constraint somewhat by moving allowances that they own from their normal "trading" allowance account, to their "compliance" account. Only covered emitters have compliance accounts. Allowances moved into these accounts cannot be traded, they may only be used for compliance at the end of some future compliance period. The exemption a covered source can get by moving allowances into a compliance account is limited by emissions from the sources most recent verified emission report.[6] So, in order for large sources to comply with the holding limit, they will need to move allowances into their compliance accounts to the maximum extent allowed. Even then, since the compliance account exemption is limited by recent past emissions, firms facing a high demand year after a couple of low demand years may have trouble legally accumulating enough emissions to comply with the law. The quarterly auction and PCR sale rules explicitly disallow any bids that would put a bidder into non-compliance with the holding limit (Van Horn, 2013).

---

[4] While banking of allowances is generally allowed, borrowing from future periods has not been. We will confine our attention to the relevant case of banking.

[5] The CARB proposal follows closely the recommendations of a white paper that it commissioned on market manipulation (Harris, 2010). For reasons that are not made entirely clear, this white paper applies to allowances the ownership limits that the Commodities Futures and Trading Commission (CFTC) applies to derivatives contracts.

[6] Van Horn (2013) provides a set of example calculations on how large emitters will need to manage their trading and compliance accounts in order to comply with both the holding limit and their compliance obligation.

Municipal electric utilities, although not subject to the same regulations as investor-owned utilities, also face an incentive for shifting allowances into their compliance account. They receive their allowances for free, but must choose between placing their grandfathered allowances in their compliance account or consigning them to the auction. It is not possible for them to hold grandfathered allowances in their trading account.

Our laboratory experiments examine the possible effects on market outcomes of the price containment reserve sale and the holding limits. These two features of the AB 32 design are linked by their relationship to market liquidity and to recurring concerns over the behavior of price in markets with a fixed supply of allowances.

In the next section, we discuss the connection between price collars and holding limits by assessing their likely effects on market outcomes, in particular through their effect on liquidity. We also discuss how particular AB 32 design features may interact with these liquidity effects. In the section that follows, we report on experiments where we test the AB 32 holding limit and price collar for their effects on liquidity, price discovery, and market efficiency. We find that, in our laboratory setting, the holding limits result in a large reduction in market liquidity, which has the unintended consequence of increasing price volatility and reducing market efficiency. Because market manipulation is known to be facilitated by periods of low liquidity (Pirrong, 2009), the AB 32 holding limit has the unintended consequence of possibly facilitating common forms of market manipulation. This points to the need to assess what levels of concentration may give larger market participants damaging price-setting power so that holding limits, if present, can be only as binding as is necessary. Future research should address whether there are less costly mechanisms for limiting market power of large emitters. Improved price collar designs may help.

Our experiments also demonstrate how the three tier price containment sale, by providing a significant, albeit expensive, additional supply of allowances, can give market participants a buffer against the worst consequences of low liquidity periods. Buying from the reserve in anticipation of future scarcity adds a small element of "borrowing" from future supplies, since theses reserve allowances have no vintage and may be used in any period once purchased. These results justify increasing attention to the design of price collar mechanisms for improving the performance of allowance markets.

1.3. Liquidity, price collars, and market manipulation

A good with a truly zero elasticity of supply seems a most unnatural creature. The reason for this is well-understood. As the price of such a good rises, the incentive to add to the supply becomes irresistible. And yet, the early models of emission trading schemes constrained bureaucrats to choose a single value for emissions rather than a supply schedule.[7] The difficulties with this assumption did not escape researchers' notice for long. It soon became apparent that a fixed vertical supply curve lead to excess price volatility (Weitzman, 1974) and, though maybe it should not have been a surprise, opportunities for market manipulation (Hahn, 1984), the profitability of which generally depends on supply frictions of one sort or another (Hart, 1977; Jarrow, 1992; Pirrong, 2009). The unnatural price volatility can lead to a variety of resource misallocations as well.

[7] Presumably this was influenced by the structure of then existing command and control regulations that operated as a fixed standard of one sort or another.

While it was only two years after Weitzman's article, that Roberts and Spence (1976) noted that relaxing the slope of the supply curve could limit price volatility while retaining the informational advantages of market exchange, more than 40 years after the Roberts and Spence article, economists still have not fully resolved the question of how to build responsiveness into the supply of emission allowances, probably because there is no single optimal strategy, everything depends on the specific context. A large recent literature has explored the effects of adding responsiveness to the supply curve with a "price collar", or a combination of price floor, below which no allowances are sold by the regulator, and a price ceiling, above which the supply is expanded (Dinan, 2010).

A parallel literature on the competitiveness of fixed supply allowance markets has also grown quite large. With the recent discussion of a national cap and trade program on GHG emissions, considerable attention was paid to what sort of market regulation might be required to mitigate the obvious opportunities for market manipulation that arise when firms face criminal penalties for not obtaining their needed share of a fixed pot of allowances (Monast, Anda, & Profeta, 2009; Pirrong, 2009). Since allowances are essentially an input to production of some output, manipulation of the allowance market might be used to gain competitive advantage in the output market (Goeree, Palmer, Holt, Shobe, & Burtraw, 2010; Montero, 2009). Goeree et. al (2010) describe laboratory experiments where those receiving large grandfathered allocations withheld supply of allowances to the market thereby raising the price of permits and helping maintain higher prices in the output market.

One feature of existing emission markets that has helped soften the supply constraint has been banking. For early implementations of cap and trade schemes, initial over-allocation of allowances along with liberal rules on banking, lead to more concern with prices being too low rather than too high. Large stocks of allowances relative to the compliance need did not provide a fertile environment for price manipulation. A recent summary of empirical work on manipulation in emission markets finds little convincing evidence for significant manipulation in these markets to date (Montero, 2009). It could be that this was a lucky break deriving from over-allocation in markets where banking was allowed, leaving the early cap and trade markets awash in liquidity, which is the first defense against manipulation.

Price containment mechanisms are closely related to market liquidity. The hard price cap and price floor first proposed by Roberts and Spence provides unlimited liquidity at the price cap and zero liquidity below the price floor. Auctioning of allowances as opposed to giving them away for free probably has the effect of greatly expanding liquidity. Allowances sold at auction are, by definition part of the float on the day of the auction, but the same will certainly not be true of grandfathered allowances. In particular, it is not well understood how available to the market are the allowances allocated to other participants including municipal electric and gas utilities and industrial sources. What is clear it that grandfathering without requiring consignment can only lower liquidity relative to auctioned allowances.

2. Research questions

2.1. The efficient equilibrium

With the emissions cap, and hence the number of allowances available from the state, declining over time, theory suggests that, other things equal, traders will add to the bank of allowances set aside for future use whenever the expected increase in allowance value is greater than the rate of

interest. In equilibrium, then, the price of allowances must be rising at the rate of interest whenever there is a positive stock of banked allowances.

The California energy market is subject to large fluctuations in the demand for allowances due to random variability in the supply of non-fossil fuel generation at hydroelectric facilities and nuclear power plants. Firms anticipating these future periods of high allowance demand would be expected to accumulate extra allowances during low demand years to lower their exposure to price spikes in these "low hydro" years. Again, firms will compare additional investments in allowances to their expected rate of return on other assets. Allowances represent a risky asset which traders will purchase so long as the risk-adjusted return is greater than that available on those other assets. In equilibrium, these rates of return should be equal.

In our experimental setup, we use a zero interest rate for simplicity. The only reasons participants would choose to bank allowances from one period to the next would be the expected future scarcity due to the tightening of the cap and the allowance price risk due to exogenous shocks, which we will model as periods of high allowance value induced by random shocks to output demand. We will refer to these periods as 'low hydro' years because the unavailability of hydroelectricity due to random periodic droughts. High allowance demand could also be caused by an exceptionally hot summer with accompanying demand for air conditioning, or by an extended, unplanned outage at a nuclear power plant.

In a market with full information and perfect foresight, we would expect the market price to equal the long-run marginal abatement cost. This implies that the market price of allowances would be constant across all periods. Any deviation of the market price from the long-run marginal abatement cost would result in the opportunity for profitable trades, which would push the price back down to its dynamically efficient level. Assuming further that the cap was set optimally, this implies that the market price is also equal to the social opportunity of the use of the allowance for emitting greenhouse gases. This perfect foresight dynamic equilibrium would maximize the net social product in the market. A "myopic equilibrium", where there is no banking, would see prices rise over time and spike up in periods of high demand.

In our laboratory setting, we expect that subjects will participate in strategies to level their costs as the cap declines and to insure against random periods of high demand. With zero interest rates and no restrictions on banking, this implies positive levels of banking and the shift of market prices away from the myopic path that would occur if there were no ability to shift allowances between periods. Banking should be associated with a price that stays closer to the long-term marginal abatement cost and that rise less during periods of high demand, when allowances are particularly valuable in production.

2.2. Market manipulation

The manipulation of the allowance market by a participant acting non-competitively may drive a market away from the efficient equilibrium. Market manipulation has been observed and widely studied by economists. Manipulation can be made profitable by frictions in exchange or information flow, or by a participant having control over a sufficient share of the available liquid supply of allowances that it becomes possible to strategically manipulate the price. Either way, market manipulation depends on low levels of 'float' or available liquidity relative to the control over assets that can be achieved by a small group of participants.

Although there is little substantial evidence of manipulation in emission allowance markets to date (Montero, 2009), following recommendations in a report prepared for the Western

Climate Initiative (Harris, 2010),[8] CARB chose to implement a limit on ownership of allowances. The limits on allowance ownership recommended by Harris and subsequently approved by CARB were copied from limits that the CFTC applies to regulated derivatives.[9] Harris notes that the CFTC limits ownership of derivatives of agricultural commodities to 10 percent of prior year's open interest up to 25,000 plus 2.5 percent of anything above that amount. He then uses this example to recommend that CARB limit ownership of actual allowances (as opposed to derivatives contracts) of 2.5 million metric tons plus 2.5 percent of the total allowance budget for the year. This rule is recommended not for its relevance to preventing market dominance by a single large owner of allowances, but by analogy to manipulation that uses derivatives markets to take advantage of periods of low liquidity in related commodities markets.[10]

In our experimental investigation, we do not address the question of whether these ownership limits are tighter than they need to be to prevent market dominance. Rather, we will assess their likely effect on measures of market outcomes to examine whether the ownership limit, if it is binding, reduces the gains from trade.

Initial analysis by Van Horn (2013) argues that the ownership limit will be sufficiently tight that, under reasonable scenarios, the larger California sources may be uncomfortably close to being unable to purchase the allowances needed to meet compliance obligations. This rule will likely result in larger sources moving large blocks of allowances from their trading accounts into their compliance accounts just in order to meet compliance obligations. Movement of allowances from trading accounts into compliance accounts can also be expected of municipal utilities not near the holding limits due to their inability to bank grandfathered allowances in their trading accounts. As a result, limitations on the ability of larger holders to accumulate allowances in advance of higher demand years may be expected to add to price volatility and to reduce the inter-temporal efficiency of the allowance market. The incentive that covered entities have to move allowances from their trading accounts into non-tradable compliance accounts, should also be expected to lower liquidity in the market.

2.3. Price containment reserve

The price containment reserve, although generally intended as a mechanism for reducing price volatility in these markets that may exhibit very inelastic supply, has the added property of providing liquidity to the allowance market, albeit possibly at a fairly steep price. This means that the price containment reserve might operate in the opposite direction to the ownership limit, increasing liquidity and improving opportunities for price-smoothing arbitrage.

Two different models of price containment reserve have been implemented in emission allowance markets to date. The California, three tier, price containment reserve sale, as already described, operates something like a discriminatory price auction with only three allowable bid levels. A different model has been chosen for RGGI, where the reserve is released as part of the

[8] Current membership in the Western Climate Initiative includes California, British Columbia, Manitoba, Ontario, and Quebec.

[9] At the time of this writing, that rule had been stayed by the U.S. District Court for the District of Columbia.

[10] In particular, Harris does not provide any evidence that limiting allowance ownership to less than 1/20th of the amount released in one year by the state is necessary to avoid price-setting behavior by a dominant trader.

auction reconciliation process. If the auction would close above a reserve trigger price, then some or all of the allowances in the reserve are made available at the trigger price in order to keep the price from rising above that price.[11] Should the reserve be exhausted, the price could continue to rise if necessary to allocate the available allowances to the highest bids.[12]

At first glance, these two reserve sale mechanisms appear to provide equivalent flexibility of supply; each adds the same level of liquidity at the same trigger price. The question we examine in our experiments is whether the different types of sale lead to the same pattern of allowance release and whether any differences in the pattern of release induce changes in measures of market performance. Although the total supply of allowances is identical in the two cases,[13] allowances in the California post-auction sale are available for use in inter-temporal arbitrage while the RGGI reserve allowances are not. In fact, the availability of advance purchase of allowances from a reserve that would otherwise only be tapped in years of high prices, provides an opportunity for a form of limited "borrowing" of allowances from future years.

These different price containment sale mechanisms also imply a different ability on the part of traders to address the risk of future scarcity. A trader who thought that there was a chance that a given tier of allowances might be exhausted by a certain time in the future, may wish to buy from that tier in advance rather than risk not having any allowances in that tier during a high demand period in the future. In an allowance market where the regulated participants are risk averse to being found in non-compliance, traders may buy down the reserve in advance, leaving less available during years of high demand. It is important to know whether the availability of advance purchases changes the performance of the reserve and how any differences observed affect market performance.

The interaction of the holding limit and the price containment reserve through their effect on market liquidity suggests that we jointly investigate the effects of holding limits and the price containment reserve sale. We wish to test whether tight holding limits cause significant deterioration in measures of liquidity and of market performance, and how holding limits interact with the design of the price containment reserve sale. We use a laboratory set-up that mimics the California market design in many important respects and that also mimics the anticipated variability of demand for emission allowances.

3. Method

3.1. Experimental setup

Our laboratory sessions were implemented with the University of Virginia's auction software (Veconlab) and were designed to mimic key features of the California allowance market,

[11] There are other differences besides those described here. For example, the California reserve, as initially implemented, is a one-time pot of allowances that, once exhausted, never refills. The RGGI reserve refills each year. In our experimental sessions, both treatments use a fixed, one-time reserve.

[12] An often discussed variant of the price containment reserve has the top tier of reserve be unlimited, meaning that sufficient allowances would always be made available to prevent the price from rising above that point. This "hard cap" on allowance prices could be built into either the California or RGGI style reserve mechanisms.

[13] In our experiments, we use a three price-step trigger for releasing allowances into the auction, so that the supply curves are identical in the two treatments.

maintaining details that would likely affect subject responses to variables of interest: holding limits, price containment reserve, and demand spikes. Each session comprised a sequence of 12 periods (or rounds) of "permit" markets,[14] where the permits (allowances) were required for production of an "output" good that sold for a fixed price that was announced at the start of each round. Sessions took 1.5 hours, half an hour for reading the instructions and one hour for the series of 12 rounds.

Each session had 12 experienced[15] subjects, who were students at the University of Virginia. Subjects were paid a US$6 show-up fee plus their earnings during the sessions, with an average total of around US$40. All subjects had 5 "capacity units," each of which could, if operated, produce 1 unit of output per period. All subjects received a free allocation that declined after each compliance period, reflecting the gradual reduction in free allocations in the actual program. Free allocations were zero in the last three rounds. High users received an allocation twice as large as that for low users. Subjects were assigned one of four different roles, which stayed the same throughout the session. Half were "low users," requiring one allowance to operate each capacity unit, and half were "high users" requiring two allowances per capacity unit.[16] Half of the subjects in each of these groups were required to consign all of their free allocations of allowances to auction and half were not required to consign any but had the option to consign.

The number of allowances made available in each period declined steadily over time. This schedule was published in advance and available for reference by the subjects throughout the session. The allowances could be banked into future periods in either a holding account or a compliance account, with the compliance account allowances no longer available for any use other than to meet compliance obligations for that subject. "Compliance periods" were defined such that true-up occurred every third round, at which time any deficits in allowances relative to requirements were tripled.[17] Allowances had no redemption value after the end of the last period.[18]

A *period* comprised a sequence of information disclosures and decisions in which the subjects are asked to:

---

[14] The language that we used in the laboratory sessions was neutral as to the purpose of the market exercise referring to "permits" rather than "allowances," and no mention was made of emission markets.

[15] All subjects were 'experienced' in the sense that they had participated in one earlier training session with somewhat different settings in order to familiarize them with the rather complicated setup.

[16] One can think of this as reflecting the difference between, say, natural gas and coal-fired generators.

[17] In the last period of a session, the penalty was calculated as three times the closing price of the auction in that period and was subtracted from earnings.

[18] This may result in some differences between what we observed in our experiments and what we would have observed in sessions with many more rounds, due to permits losing their value at the end of the 12th period. There is no easy way around this endpoint dilemma, but it may not be as significant as it first seems since the same thing is true in all treatments. So for this effect to be important, it would have to have a substantially stronger effect in one treatment than the other. There is nothing about our results that indicates such an asymmetric effect.

1. Observed the price of the output for that period, any must-run requirements, and one's own current trading and compliance account totals;
2. Chose how many allowances to consign to the auction in addition to any required amounts;
3. Chose how many allowances to move from their trading account to their compliance account;
4. Submitted bids in auction for any number up to the current purchase limit and view auction results, including the release of any permits from the in-auction containment reserve for sessions with such reserves;
5. Posted offers to buy and sell in the post-auction spot market and view own spot market results;
6. Posted offers to buy in price containment reserve sale (if the reserve was not exhausted) and view results of the sale; [Note: This step only occurred in treatments with a separate price containment reserve sale.]
7. Decided how many units to run above any must-run requirement;
8. Observed earnings, account totals, cash balances, and any deficits relative to compliance requirements.

The auction in each period had a uniform-price, sealed-bid format. The number of allowances for sale included a base auction quantity, which declined each period, plus any allowances consigned to the auction. The number of bids that a subject could submit was determined by the purchase limit or the holding limit. A bid that would satisfy the purchase limit for a subject was not allowed if winning that bid would result in a violation of that subject's holding limit. Bidders with winning bids paid only the closing price of the auction for any accepted bids.[19]

The spot market, which followed immediately after the auction, was a single-round, limit-order call market where each subject could post one buy order and one sell order. A buy order consisted of a quantity offered and a maximum price to be paid, and a sell order was a quantity along with a minimum acceptable price. The bids and asks were ordered to determine the market clearing price. All bid and ask orders were filled at that spot market price, and bidders were informed of any changes in their cash and permit holding accounts.[20]

The costs of operating capacity units were randomly assigned each period and were uniformly distributed in the range of $4–20 per unit for high emitters and $14–22 for low emitters. The maximum value to the subject of a permit for a given capacity unit is equal to the output price minus the cost of operating the unit, all divided by the number of allowances required to operate the unit.

For all sessions used in the analysis, there was a two-thirds probability of a $35 output price and a one-third probability of an output price of $60 (or in some cases $50). The higher output price induces a high willingness to pay for allowances and mimics periods of high emissions with accompanying high allowance demand, such as might occur in a drought year with low

[19] Half of the sessions used the first rejected bid to set the price, the rule in the RGGI auction, and half used the lowest accepted bid, which is the rule in the California auction. No differences in outcomes between these closing rules were observed.

[20] Ties are determined by a random device.

availability of hydroelectric power.[21] During "normal" periods, low users were required to run 3 of their 5 units while high users had no "must-run" requirement. During "low hydro" periods, the must-run requirement for low users rose to 4 and high users had a must-run requirement of 1 unit.

Subjects earned a cash payment equal to 5 percent of their nominal earnings in the experiment. This amounted to approximately $50 in average earnings per subject for a 1.5 hour session.

3.2. Treatments

Loose vs. tight holding limits: Each session had either a tight holding limit of 7 for low emitters and 14 for high emitters, an amount just greater than the auction purchase limit for each subject; or a loose holding limit of 13 and 26 for low and high emitters respectively. The tight holding limit, at just greater than the purchase limit, was still greater than any routine contemporary compliance needs. Loose holding limits allowed subjects to hold more than twice as many permits as needed to cover maximum possible production of 5 during any period. The holding limit was the treatment variable for the first set of sessions. Since we were examining the effect of tight holding limits, the treatment with the loose holding limits was our baseline treatment, and the tight holding limit case was the alternative treatment.

Post auction vs. in auction sale: A one time (non-replenished) allowance reserve was established at the start of each session. The reserve was divided into three price tiers of one third of the available reserve each, and once an allowance was allocated to a price tier, it remained in that tier until sold or until the end of the session. Allowances in these tiers were sold at $23, $28, and $33 respectively. These allowances were released in one of two ways, each corresponding to an experimental treatment: a California-style, three-tier sale that takes place after the auction or a RGGI-style release of allowances into the auction itself at the same prices as in the three-tier sale. For the sessions with a post-auction price containment reserve sale, orders were filled from the price tier in which the bid was made unless the "pull-down" rule provided that the order be filled at the next lower price. Oversubscribed tiers were filled in proportion to the bidder's share of the total number of bids at that tier. All of the sessions used for testing the effects of the price containment reserve treatments on market outcomes were sessions with tight holding limits.

3.3. Sessions

Our experimental sessions are listed in Table 1. A preliminary round of sessions, differs from the later sessions in two key respects, they had a somewhat less binding cap on emissions, a lower output price in low hydro periods, and, unlike the other sessions, the allowances in the price containment reserve were reallocated at the end of each round so that each tier always had one third of the remaining allowance left in the reserve. This tier reallocation results in a different pattern of purchase from the reserve, as we will discuss later. The results from these sessions differ in important ways from the remaining sessions and so are not strictly comparable. We will show how the exclusion of these sessions affects our hypothesis tests.

A second group of sessions, are used for our main tests of the effects of tight versus loose holding limits. These sessions all use the California-style post-auction price containment reserve

[21] As we discuss shortly, a preliminary set of sessions with a relatively loose cap were run initially. These sessions had other differences which limit their use in some hypothesis tests. But these sessions are informative and we note when we use them to support our analysis.

sale. The third and final group of sessions are used for testing the post-auction reserve sale against the RGGI-style of in auction allowance release. For this comparison, we use the tight holding limit sessions from the second session group along with four additional sessions that differ from the other four sessions only in the reserve sale treatment. Half of the sessions in these last two blocks use the highest rejected bid for determining the auction closing price and half use the lowest accepted bid.

Table 1: Experimental session summary information

Experimental sessions

Session	Tiers	Holding limit	Reserve sale	Auction price rule	Cap start - end	Reserve quantity	Output prices	Total session cap	Dynamic efficient price	Maximum surplus	Block 1	Block 2	Block 3
1	Reallocated	Loose	Post auction	Highest rejected bid	63 - 30	24	35, 50	558	$ 15.00	3441			
2	Reallocated	Loose	Post auction	Highest rejected bid	63 - 30	24	35, 50	558	$ 15.00	3441			
3	Reallocated	Tight	Post auction	Highest rejected bid	63 - 30	24	35, 50	558	$ 15.00	3441			
4	Reallocated	Tight	Post auction	Highest rejected bid	63 - 30	24	35, 50	558	$ 15.00	3441			
5	Reallocated	Tight	Post auction	Highest rejected bid	63 - 30	24	35, 50	558	$ 15.00	3441			
6	Reallocated	Loose	Post auction	Highest rejected bid	63 - 30	24	35, 50	558	$ 15.00	3441			
7	Fixed	Loose	Post auction	Highest rejected bid	69 - 14	36	35, 60	498	$ 14.75	4727			
8	Fixed	Tight	Post auction	Highest rejected bid	69 - 14	36	35, 60	498	$ 14.75	4727			
9	Fixed	Loose	Post auction	Highest rejected bid	69 - 14	36	35, 60	498	$ 14.75	4727			
10	Fixed	Tight	Post auction	Highest rejected bid	69 - 14	36	35, 60	498	$ 14.75	4727			
11	Fixed	Loose	Post auction	Lowest accepted bid	69 - 14	36	35, 60	498	$ 14.75	4727			
12	Fixed	Tight	Post auction	Lowest accepted bid	69 - 14	36	35, 60	498	$ 14.75	4727			
13	Fixed	Loose	Post auction	Lowest accepted bid	69 - 14	36	35, 60	498	$ 14.75	4727			
14	Fixed	Tight	Post auction	Lowest accepted bid	69 - 14	36	35, 60	498	$ 14.75	4727			
15	Fixed	Tight	In auction	Highest rejected bid	69 - 14	36	35, 60	498	$ 14.75	4727			
16	Fixed	Tight	In auction	Highest rejected bid	69 - 14	36	35, 60	498	$ 14.75	4727			
17	Fixed	Tight	In auction	Lowest accepted bid	69 - 14	36	35, 60	498	$ 14.75	4727			
18	Fixed	Tight	In auction	Lowest accepted bid	69 - 14	36	35, 60	498	$ 14.75	4727			

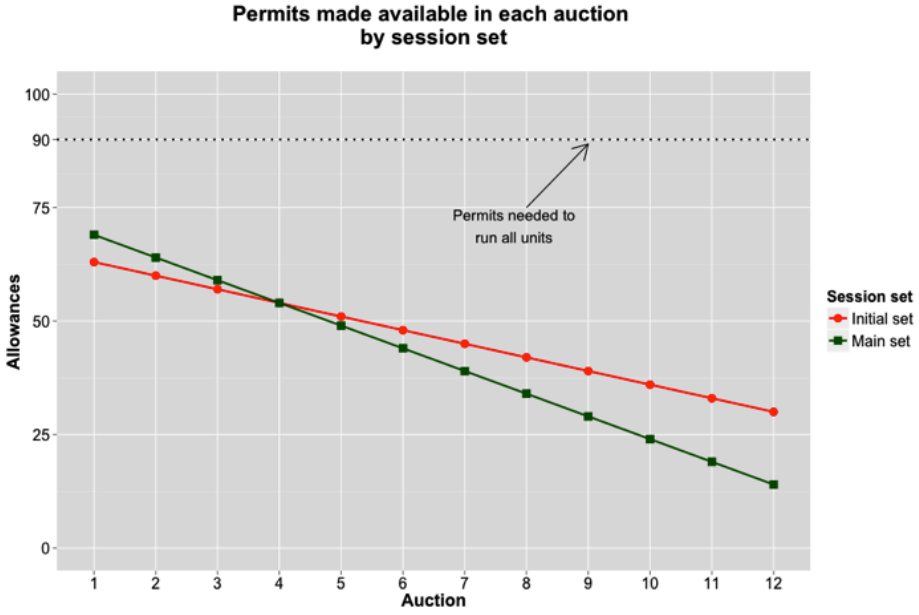


Figure 1: Emission cap by period

3.4. Experimental Measures

3.4.1. Liquidity

Market manipulation, if successful, imposes real costs on society. This is not merely a theoretical concern. Of particular relevance, the California energy market suffered a very damaging episode of manipulation in 2001. While the details of each instance of market manipulation, including the case of the California energy market, are different, studies of market manipulation have reached a strong consensus that improving the liquidity of a market is the best defense against manipulation (Hart, 1977; Jarrow, 1992; Pirrong, 2009; Putniņš, 2012). Liquidity is a measure of how many units of an asset (in our case allowances) are available to come into the market as the price rises. The greater this amount relative to the size of the market, the greater is market liquidity. It not only makes profiting from price manipulation more difficult and more risky, but also improves the overall competitiveness of the market, which can be expected to enhance price discovery. In light of this definition, a dominant position in a market may be seen as a particular form of low liquidity; there are few allowances for sale except by the owner in the dominant position.

Market liquidity is one important measure of market performance. Market manipulation generally depends on making a purchase or a sale that would likely not be profitable were it not for the effect on market prices. Our experimental sessions have four sources of liquidity, the number of allowances provided at auction by the state, allowances held in trading accounts, allowances consigned to the auction, and allowances offered in the spot market. The last three of these may only come from allowances held in subject trading accounts. The base auction amount is identical across treatments.[22] We used the number of allowances in trading accounts as a key measure of liquidity and also looked at what we can directly observe about the float.

The holding limit provisions of the California allowance market are unique among emission trading programs and have been the subject of considerable discussion (Linklaters, 2011). One of the key purposes of this study is to examine the effect of the holding limit provision on the liquidity and overall performance of the allowance market. The results of our investigation can help shed light on the issue of market manipulation and on the use of the holding limit to avoid it. If it is true that, contrary to expectations, holding limits may result in increased risk of certain kinds of manipulation, then there should be an effort to determine whether holding limits are a necessary and effective tool for preventing market dominance.

3.4.2. Price discovery and price volatility

It is a well-established result that the price of emission allowances should be close to the "marginal cost of abatement." In our laboratory setup, an individual subject can only abate emissions by reducing production. So, in aggregate, abatement is accomplished either by a reduction in production or by a shift of production from high emitters to low emitters. In theory, the price of allowances will equal the value of the permits to the last unit that is not run in order

[22] Some commentators on the California market have noted that one straightforward way to increase liquidity would be to increase the frequency of auctions. The RGGI program settled on quarterly auctions largely due to the uncertainty over the administrative burden that the auctions would represent. As experience with allowance auctions has grown, the rationale for low frequency auctions has faded. Increasing auction frequency seems to provide a low cost way of greatly enhancing market liquidity.

to keep emissions under the cap. Any other price would leave the sources with a profitable trade, and result in higher than necessary costs of compliance. It follows that we should use the proximity of the market price to the marginal cost of control as one standard of market performance.

Our experimental setup determines the costs of production, the output price, and the supply of allowances. Since there is not abatement technology in our setup, the value of allowances is completely determined by their use in production, and the abatement decision is the same as the decision to run or not to run a production unit. The value of allowances, then, is determined by the output price minus the cost of production, taking into account the number of allowances needed for each unit (one for "low users" and two for "high users". Since the cost of storing allowances from one period to the next is zero, then absent risk, we can calculate which unit is the marginal unit over all periods in a session. The value of allowances in production for this unit is the dynamically efficient price of emissions. If one market design element results in market prices that are systematically further away from the marginal abatement cost, then that design element harms the market's price discovery function. We use the mean absolute difference between the market price and the dynamically efficient price as a key measure of market performance.[23]

Uncertainty adds to the cost of doing business. Added volatility in allowance prices, other things equal, tends to increase the cost of complying with emission reduction requirements. The very presence of the price containment reserve designs we test resulted from concerns over excess volatility. Price volatility, then, should be one of our metrics of market performance. We measure price volatility as the variability of auction closing prices.

3.4.3. Efficiency

Efficiency is closely related to price discovery. As the price moves away from the marginal abatement cost benchmark, businesses receive incorrect signals about which units of abatement are worthwhile and which ones are not. This results in cases where some production units worth running do not run and some units not worth running do run. In both cases, society loses valuable resources either as costs that are higher than they should be or benefits that are lower than they should be. In the absence of the need for allowances, efficiency would be measured by the difference between the cost of production and the output price. In an emissions market that is not a correct measure of surplus. The restriction on emissions implies that the emissions have a social cost. In theory, an efficient cap on emissions will result in an emissions price equal to the social cost of the last unit of emissions and that all infra-marginal emission reductions would have a value at least as high. For our measure of efficiency, we value each unit of emissions at the dynamic optimal price. This amount must be subtracted off of private surplus of operating the unit that results in the emissions. We measure efficiency as the net value to society of a given unit of production; that is, the price of the output minus the cost of operation minus the cost of GHG emissions.[24] This efficiency measure implies that, other things equal, allowances taken from the price containment reserve would lower the social surplus if all of the other allowances were applied to the most efficient production units. Allowances charged as penalties and retired

[23] Another possible measure of difference is the mean absolute deviation. This measure leads to the same conclusions in our experimental tests.

[24] The social cost of GHG emissions is assumed to be the dynamically efficient price, which is the marginal cost of reductions for a session.

would reduce otherwise optimal production away from the efficient equilibrium. The actual effect of these marginal allowances on the social surplus depends on the other production decisions during the session and cannot be predicted in advance.

Table 2: Hypotheses Tested

	Null (H_0)	*Alternative (H_1)*
Liquidity	*Holding limits do not reduce market liquidity*	*Liquidity is lower under tight holding limits*
Price discovery	*Tight holding limits do not increase the difference between the observed price and the efficient market price*	*Tight holding limits increase the difference between the efficient price and the observed price.*
Efficiency	*Tight holding limits do not reduce efficiency*	*Efficiency is lower under tight holding limits*

4. Results

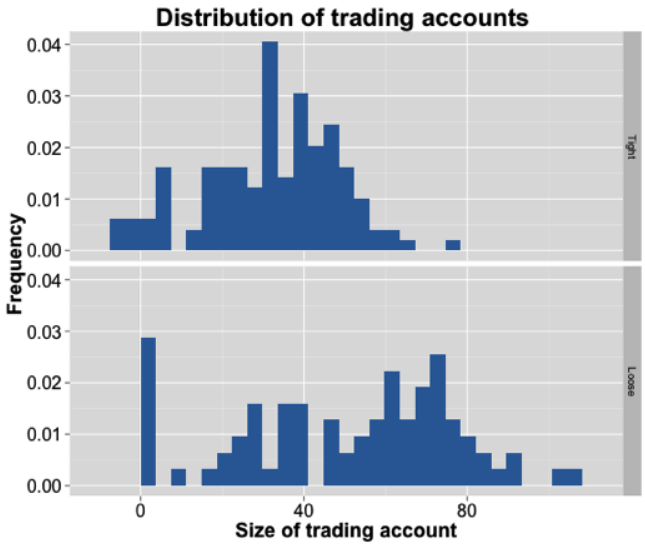
4.1. Effect of Holding Limits

4.1.1. Liquidity

Each laboratory session had either a tight holding limit or a loose holding limit, with tight holding limits being just greater than the auction purchase limit for each subject and loose holding limits being twice the auction purchase limit. The looser holding limit is still tight enough to bind on occasion, but with much lower frequency and at much higher banking levels. The holding limit is the treatment variable for this set of sessions. Given the constraint the tight holding limits place on participant ability to bank allowances in tradable accounts, and given the incentive that participants have to move permits into non-traded accounts, our expectation is that liquidity, as measured by the number of banked allowances available for trade, would fall. The effect of the tighter limit on subjects in our sessions was considerable.[25] Figure 2 shows the difference in the size of the tradable bank for each period separated by treatment. For most of the auctions, the average size of the tradable bank under the loose holding limit is double that under the tighter limit. Figure 3 shows the frequency distribution of different trading account sizes with tight holding limit observations in the upper panel and loose in the lower panel.

Figure 2: Trading account holdings by period by treatment

Figure 3: Distribution of holding account size by treatment



Of the banking that is done, trading accounts make up a much larger share of the total bank in the loose sessions as compared to the tight ones. Trading accounts are both larger and a larger share of the total bank under loose holding limits, which is illustrated in Figure 4. This is a

[25] Unless otherwise noted, all results will exclude the preliminary sessions, since they are not completely parallel to the sessions we used for our hypothesis tests.

scatterplot of the number of allowances in holding accounts by the total number of banked allowances, so it is a way of viewing holding accounts relative to non-tradable compliance accounts and to the bank as a whole. As the size of the total bank rises, we see the number of allowances in trading accounts rises as well, but there is a striking difference in the shares of allowances in holding accounts between the two treatments. The holding accounts are distinctly larger under loose holding limits. Indeed, total banking never reaches 100 allowances under tight holding limits. Even for very small banks, where subjects are unlikely to be constrained by the holding limit, the holding accounts are smaller under tight holding limits and the share of allowances in holding accounts is higher. The lines represent the line of best fit through the points for each treatment, and the shading shows the 95% confidence interval for that line.

This loss of liquidity is not made up for in other ways. Volumes of spot market trades are not significantly different across treatments. Nor is there an increase in the consignment of permits to the auction, as illustrated in Figure 5.

Given these images from the sessions that we use for testing the effect of holding limits, it will not come as a surprise that we find a very significant difference between the treatments and easily reject the null hypothesis of no effect.

Figure 4: Trading account total by banked permit total

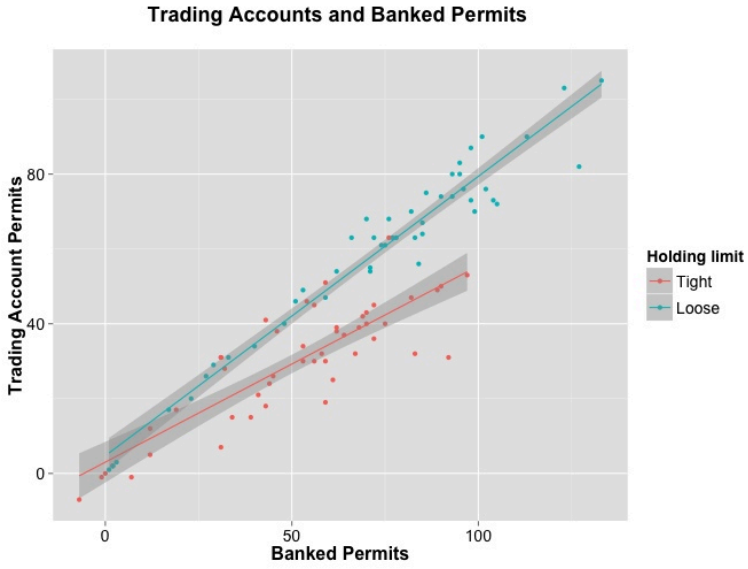


Figure 5: Consigned permits by period by treatment

Hypothesis 1: Tight holding limits will reduce the stock of tradable permits in the emission market compared to loose holding limits.

We tested the total trading account holdings under tight and loose caps using the session averages only from sessions in Block 2 (see Table 1). We also tested the hypothesis for the combined observations in session Blocks 1 and 2 using the Wilcoxon [Mann-Whitney] test stratified by session block (Hothorn, Hornik, van de Wiel, & Zeileis, 2008). With either test, we reject the null of no effect at better than 5 percent significance.

Table 3: Rank-sum tests for Hypothesis 1

Wilcoxon Rank Sum Test (two-tailed)
Session Block 2 Z = -2.309, p-value = 0.02857
Session Blocks 1 and 2 (Stratified by session block) Z = -2.2263, p-value = 0.03056

Taking advantage of the panel structure of the data (Vossler, 2013), we estimate a random effects model of trading account totals for each auction, controlling for the holding limit, whether the auction price is set using highest rejected bid or lowest accepted bid, and whether the period is a 'low hydro' (high permit price) period. The t-statistics are estimated using Arellano's estimator that is robust to both heteroskedasticity and serial correlation. The first set of estimates use only Block 2 sessions, the second set uses both Blocks 1 and 2 with a dummy variable for Block 1. The holding limit variable is highly significant as is the low hydro year variable.

Table 4: Regression results for hypothesis 1

The Effects of Holding Limits on Trading Accounts: Random Effects Regression Estimates		
	Block 2	Blocks 1 and 2
(Intercept)	19.00*** (4.01)	21.81*** (4.17)
Holding limit	28.81** (5.80)	18.42*** (3.54)
Price method	5.10 (1.03)	5.10 (0.83)
Block 1 effect		1.91 (0.32)
Low hydro year	10.71*** (9.92)	11.35*** (6.91)

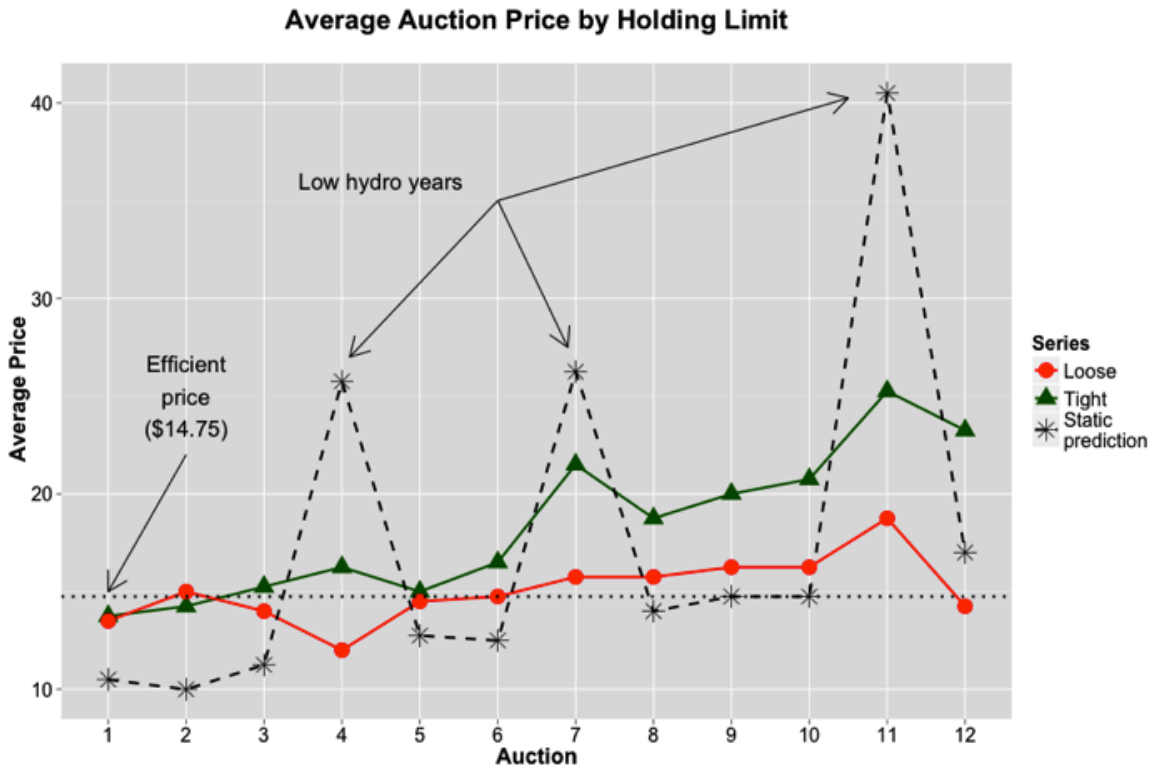
From this, we conclude that, in our laboratory setting, holding limits do indeed exert a strong influence on both the level of banking but particularly on the amount of banked permits that are available for trade. Tight holding limits will, if binding, tend to lower liquidity.

4.1.2. Price Discovery

Market liquidity is only an indirect indicator of the effectiveness of market function. Liquidity is particularly important for emission markets due to the large gains to be made by smoothing allowance scarcity and hence prices over time as allowances gradually become more scarce and as exogenous shocks shift the relative value of allowances. In this section, we focus on whether the reduced liquidity is associated with prices that either stray farther away from their socially optimal value or prices that show greater variation. Given the importance to market participants of being able to accumulate allowances based on their best assessment of current and future scarcity, we would expect that, in our experimental market, lower liquidity would be associated with prices that both stay further away from their surplus (and profit) maximizing value but that the reduced intertemporal arbitrage would lead to greater price variability as well.

Figure 6 effectively summarizes the results from the laboratory sessions. The dotted line and the dashed line represent two extreme possibilities. The dashed line gives the predicted market-clearing price in each period of a session under the assumption that subjects only take into account the current period cap, costs and output prices. It represents the "perfectly myopic" Walrasian competitive equilibrium, where subjects act as if only the current period matters and do not account for any benefits they would receive by saving allowances from one period to the next.

Figure 6: Average auction price by treatment



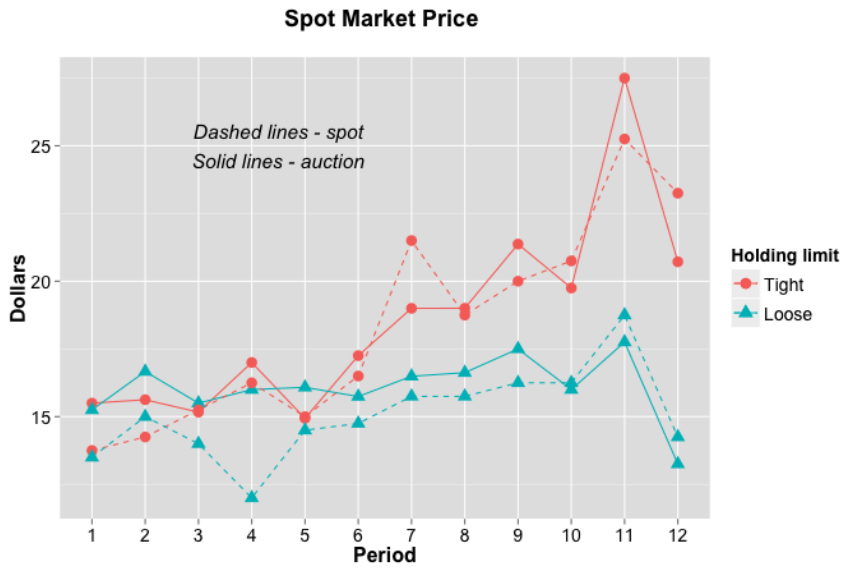
The dotted horizontal line ($14.75) shows the dynamically efficient price, which is the value of the marginal control cost combining all 12 periods in the session and represents the "perfect foresight" case, where subjects correctly anticipate all future scarcity and completely balance their allowance scarcity across all periods. If the market price exactly matched this price each period, this would maximize the value to society of the trading program because it would equate marginal emission reduction costs across both space and time, and would minimize the cost of price risk due to periodic shifts in demand.

Market participants who anticipate future scarcity can profit by accumulating allowances early and saving them for later periods, when they will have greater value as the cap declines. This would cause the observed price in the lab to be smoother and flatter than the predicted myopic clearing price. The more the market induces traders to smooth and flatten the price path away from the myopic price path toward the efficient (flat) price path, the closer the market comes to the least-cost pattern of emissions.

Smoothing the demand for allowances in this way has the salutary effect of generating earlier reductions in emissions than would occur otherwise. Anticipation of future allowance scarcity raises the true current cost of GHG emissions in the present. Incentives for early investment in new low GHG technologies rise, in turn.

The two solid lines represent the average across sessions of auction prices for each auction period separated by whether the holding limit was loose (circles) or tight (diamonds). The difference between these two lines clearly illustrates our result, that tight holding limits tend to drive the auction price further from the dynamically efficient price path, reducing the ability of traders to prepare for high demand periods, and limiting trader ability to prepare for long-term future scarcity by making additional reductions in the present and banking the unused allowances for use in later years. Price discovery is less effective with tight holding limits than with loose holding limits.

Figure 7: Spot market prices and auction prices



In considering the effects on price, our focus so far has been on the auction price. This leaves open the possibility that spot market trading serves to move prices back toward the socially efficient level. The next figure shows the average difference between auction and spot prices by holding limit treatment. While there may be some tendency *under loose holding limits* for the spot market to move price closer to the dynamically efficient price, this is not true during tight holding limit sessions.

Hypothesis 2: Tight holding limits will result in prices with a higher mean absolute deviation from the efficient market price than will loose holding limits.

As before tested the hypothesis using the session averages only from sessions in Block 2 (see Table 5). We also present results on Blocks 1 and 2 using the stratified Wilcoxen.

Table 5: Rank-sum tests for Hypothesis 2

Wilcoxon Rank Sum Test (two-tailed)
Session Block 2 Z = 2.021, p-value = 0.05714*
Session Blocks 1 and 2 (Stratified by session block) Z = 2.1319, p-value = 0.02778**

We can reject the null hypothesis that the mean absolute deviation between the actual price and the dynamically efficient price is the same in sessions with tight holding limits as in sessions with loose holding limits.

Our random effects panel regressions give similar results:

Table 6: Regression results for Hypothesis 2

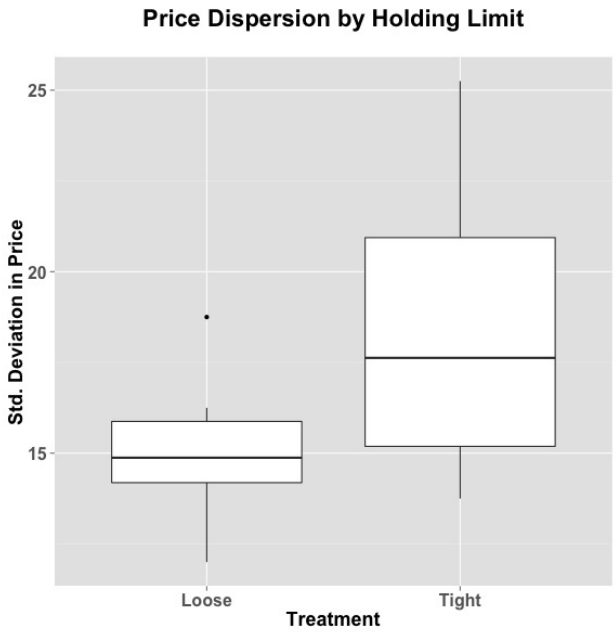
The Effects of Holding Limits on Price Discovery: Random Effects Regression Estimates		
	Block 2	Blocks 1 and 2
(Intercept)	5.08*** (7.05)	4.72*** (4.42)
Holding limit	-2.06** (-2.41)	-2.06*** (3.54)
Price method	0.23 (0.24)	-0.33 (-0.35)
Block 1 effect		0.94 (0.89)
Low hydro year	-1.96*** (-4.75)	-1.97*** (-3.45)

Price variability is closely related to price discovery. And, in fact, our result in testing for a difference in the standard deviation of price are very similar to those on price discovery, so, in the interest of brevity, we will not present them here.[26] Figure 8 shows the standard deviation of

[26] The results are available from the authors.

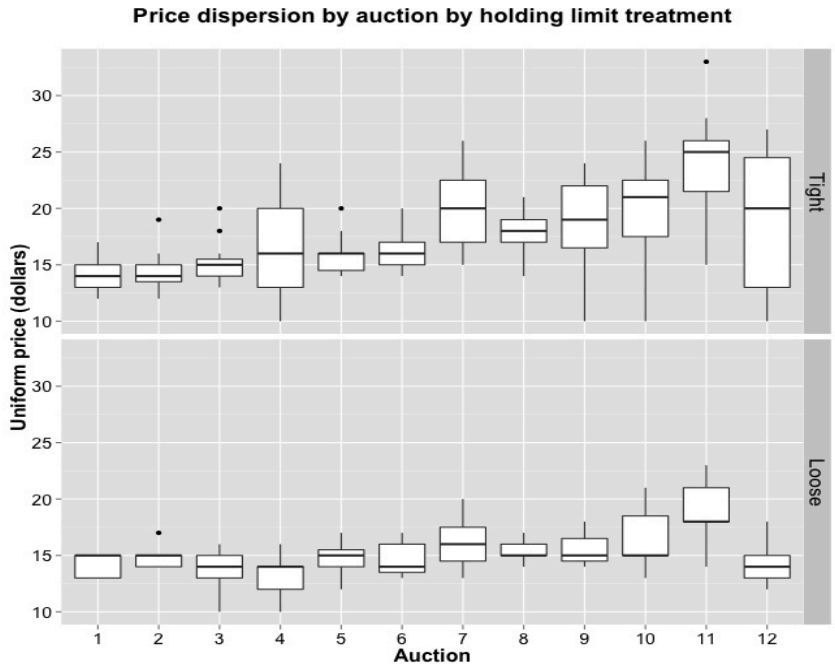
auction prices only by the holding limit treatment and Figure 9 breaks these out by auction period.[27]

Figure 8: Price dispersion by holding limit



[27] This image includes observations from Blocks 1 and 2.

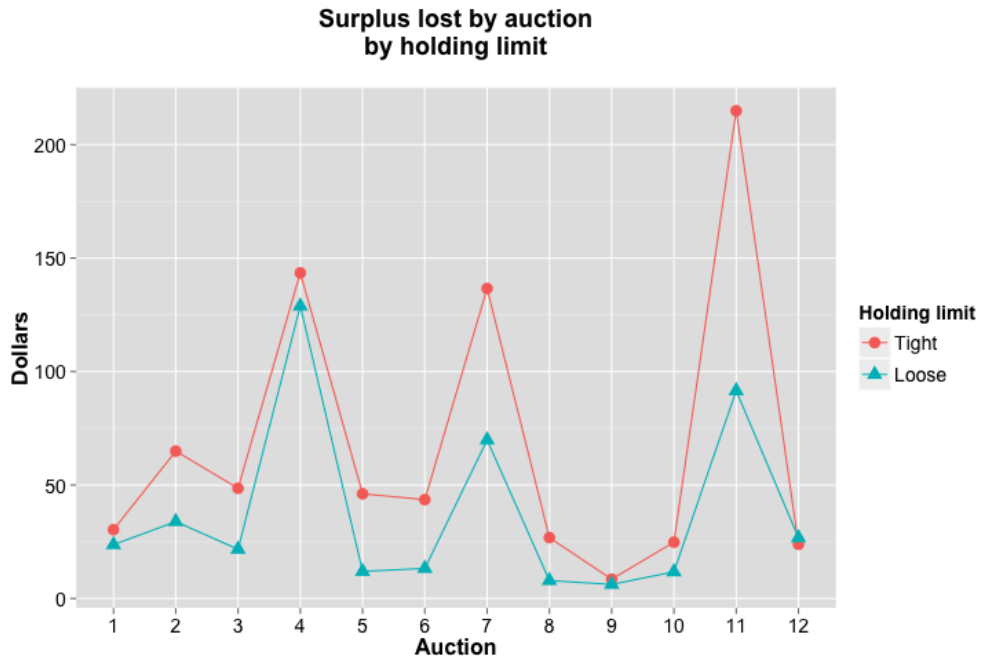
Figure 9: Price dispersion by period by holding limit



4.1.3. Efficiency

There is every reason to expect that the difficulty that lab subjects have in shifting their stock of allowances across time periods should end up reflected in lower efficiency for tight holding limits relative to loose holding limits. The permit price signals whether a particular production unit is worth running. Any movement of the realized price away from the efficient price will provide incorrect signals at the margin about running a unit. Figure 10 shows the average surplus lost in each period separated out by loose and tight holding limits.

Figure 10: Surplus lost by period by treatment



It is clear from this picture that the relatively poor price signals resulting from the tight holding limit cause reductions in surplus. The actual surplus in the Block 2 sessions, is given in Figure 11.

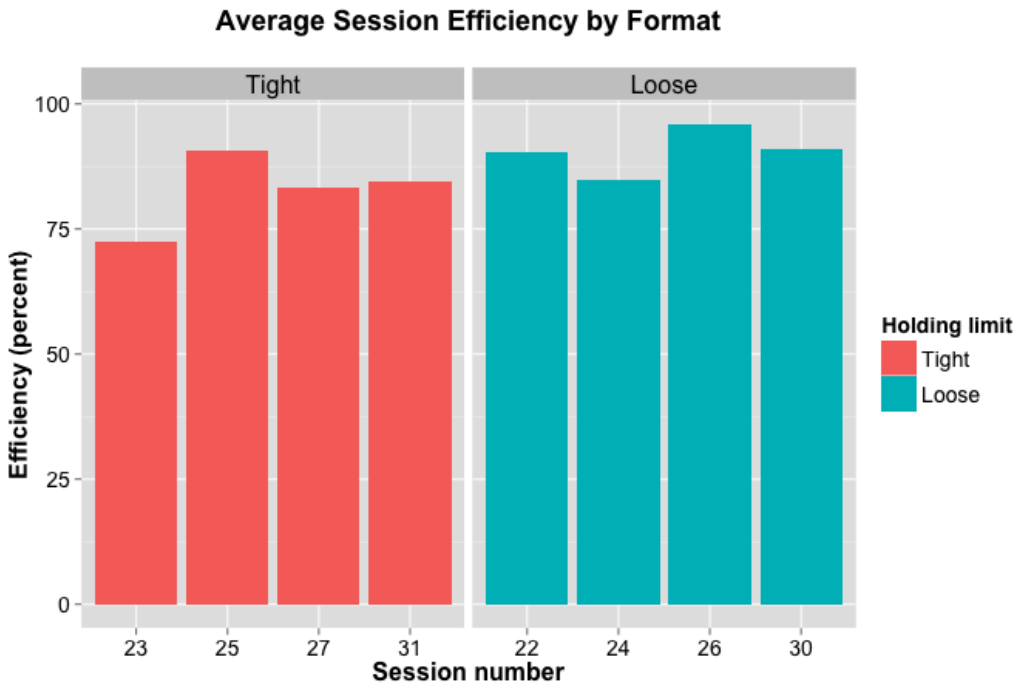


Figure 11: Efficiency by session

Hypothesis 3: Tight holding limits will reduce efficiency relative to loose holding limits.

Table 7 reports the results of the rank-sum test of Hypothesis 3. Using only Block 2 data the just misses the 10 percent level of significance, while the stratified test using both Blocks 1 and 2 allows rejection of the null at the 10 percent level.

Table 7: Rank sum tests for Hypothesis 3

Wilcoxon Rank Sum Test (two-tailed)
Session Block 2 Z = -1.7321, p-value = 0.1143
Session Blocks 1 and 2 (Stratified by session block) Z = -1.8358, p-value = 0.075*

Results from the panel regressions for Hypothesis 3 appear in Table 8. The p-values for both estimates of the effect of the tight holding limit on efficiency show significance of better than 5 percent. Although the rank-sum test on only the sessions in Block 2 just misses the 10 percent

significance level, the clear weight of evidence for these tests points to a significant effect of holding limits on efficiency.

Table 8: Regression results for Hypothesis 3

The Effects of Holding Limits on Efficiency: Random Effects Regression Estimates		
	Block 2	Blocks 1 and 2
(Intercept)	75.6***	72.5***
	(7.55)	(10.7)
Holding limit	19.8**	18.1**
	(2.31)	(-2.50)
Price method	9.71	9.71
	(1.13)	(1.13)
Block 1 effect		13.2
		(1.21)
Low hydro year	-25.8***	-38.1***
	(-4.55)	(-6.02)

These results give strong weight to the argument that tight holding limits may impose considerable costs by interfering with the efficient allocation of allowances between sources and across time. Firms in the market face higher risk and receive less accurate signals concerning where resources should be invested in emission reduction activities. Since one of the justifications for the holding limits is to avoid market manipulation, it is especially noteworthy that tight holding limits, by lowering liquidity, tend to operate directly counter to one of the key defenses against market manipulation. We conclude that tight holding limits may make the allowance market *more* vulnerable to some of the commonly discussed strategies for market manipulation. What we do not yet know is whether any protections provided by limiting the share of allowances owned by one or a few large providers outweighs the costs and risks imposed by the holding limits themselves. In the least, it does seem clear that the holding limits should be not tighter than necessary.

4.2. The price containment reserve

A price containment reserve is a stock of allowances that only becomes available for use in the market once some high price trigger-point is reached. Once the triggering event occurs, some allowances will enter the market. RGGI, for example, has chosen a reserve that becomes part of the supply of allowances available at auction whenever the auction would otherwise close above a trigger price. The allowances available at that price point would be released to accept additional bids until the supply of allowances available at that price is exhausted or the auction closes, whichever comes first.

California's approach establishes a post-auction sale where potential buyers may bid for allowances at any of three price points. The reserve allowances are divided into thirds and allocated to these "price shelves." Once an allowance is sold from a shelf, that allowance is not replaced. As long as there are plenty of allowances on a given shelf, then a bidder can make a bid at a higher shelf and expect their order to be filled from the next lower price shelf. But when a shelf gets close to empty, bidders have incentive to bid for as many as are available at that price and not to bid at a higher price because bids at the lower price are filled in proportion to the total number of bids at that price. And, bids at the higher price will likely be filled at that price and not drop to the next lower level. As already noted, our Block 1 sessions reallocated allowances

among the tiers so that, at the start of each period, all tiers had roughly equal. Our Block 2 sessions use the California rule of not reallocating between tiers.

Because of the different release mechanism and the somewhat complicated incentive structure in the California reserve, it is not clear how participants in the California market will respond to the presence of the reserve or what would be its likely effect on price. We tested whether the different PCR designs produce significantly different permit market outcomes. While our results show that there are sharp differences in the way that the different reserves were used by the subjects in our experiments, it turns out that these differences do not translate into measurable differences in liquidity, price, price volatility or efficiency. This was true for our rank-sum tests on session-level data and for our subsequent panel regressions.

For these tests, we used only observations from our Block 3 sessions. These sessions all had tight holding limits, but four used the post auction sale and four released permits into the auction when the price would otherwise close higher than the trigger. Both treatments had three price tiers of $23, $28, and $33. The reserve contained 36 permits or just over 7 percent of the total cap of 498 for the 12 period session. The rank-sum test results appear in Table 9.

Table 9: Rank-sum tests for price containment reserve effects

Rank-sum Tests for Differences between PCR Designs		
Liquidity	Price	Efficiency
Z = -0.2887 p-value = 0.8857	Z = 0 p-value = 1	Z = -0.2887 p-value = 0.8857

Since the regression results were uniformly without significance except for the intercept and the "Low hydro" variable, we do not present them here.

There are many possible explanations for the invariance of market outcomes to the price containment reserve treatment. One simple explanation is that the aggregate supply of allowances does not change as between the two and that it is the aggregate supply that matters rather than the way the allowances are given out. By this explanation, participants would be expected to use allowances the same way regardless of when they are made available. This explanation does not appear consistent with our earlier result that, during tight holding limit sessions, participants are clearly constrained in the extent to which they can time-shift the available supply of allowances and that this has a substantial influence on the path of prices and production.

Other possible explanations are that the reserve is just not large enough to make a detectable difference or that the prices are set so high that not enough of the reserve is realistically available to the market. Based on the information we have, we are not in a position to explain why the reserves as we modeled them in our experiments did not have a noticeable affect on market outcomes. We can, however, point to some differences in the way the reserves are used. These differences may point toward future explorations concerning the design of price containment reserve mechanisms.

From a participant's point of view, there may be advantages to having an opportunity to choose when to purchase allowances from the reserve as opposed to only purchasing them when some overall market trigger-point is passed, even if the individual would expect to use the allowances at the same time. Allowances that can be purchased by individual choice (even at a

high price) provide a kind of insurance to the individual, insurance that is not provided by a general relaxation of the fixed market supply curve. An individual participant who expects increased scarcity in the future knows that auction and spot market prices can be expected to be high during those periods, and there is risk that bids on permits essential for compliance will be rejected should auction prices spike more than expected. With the post auction reserve sale, individuals may see it as profitable to purchase allowances in advance at the fixed, if high, price available in one of the price tiers. The more likely it is that some period will have a price spike that will rise above the tier price, the more profitable advance purchase will seem.

Based on this perspective, we would expect a certain pattern of reserve purchases. First, we should observe purchases from the lower price tiers well before the market price reaches the trigger price for that tier. Second, we should observe more of these anticipatory purchases in tight holding limit treatments than in loose holding limit treatments, because the participants should expect greater shortages and higher spikes in price. Third, if allowances are reallocated between tiers to keep the tiers equally populated across rounds (as was true in our Block 1 sessions), we should observe more use of the reserve in advance

From the point of view of the policy maker managing the program, the obvious risk in this anticipatory purchasing is that the reserve may be drawn down in advance of need to a point where it will not serve to reduce price spikes during periods of high demand.

In Figure 12, we show the pattern of purchases from the post auction price containment reserve sale in Block 2 of our experiments. The top panel shows purchases under a loose holding limit and the bottom panel a tight holding limit. One would expect to see relatively large purchases during the high demand periods that we label as "low hydro." These high demand periods, 4,7, and 11, do show modest purchases, as expected in both tight and loose treatments.

Figure 12: Three tier PCR sales by holding limit treatment

But just as many permits were purchased in the first three periods from the $23 tier of the reserve, when the auction closing price was very close to the reserve price of $10. This appears

to be speculative buying in anticipation of substantial future price increases. These buyers are correctly anticipating that prices will eventually rise above the $23 trigger price. We observe with interest that the anticipatory buying is larger on average in the loose holding limit sessions than in the tight ones. This seems quite reasonable. In the loose holding limit sessions, these allowances may be held in trading accounts rather than having to go into the non-tradable compliance accounts. It is also interesting to note the number of permits purchased from the reserve in the years *after* a low hydro period as participants in the tight holding limit sessions restock depleted banks of permits.

There is a clear interaction between the holding limit requirement, the tightness of the holding limit, and the likelihood that firms will buy allowances out of the reserve. If this incentive is strong enough, the reserve could become depleted, as occurred in one of our tight holding limit sessions when the entire reserve was used up before the end of the session. Participants in tight holding limit sessions tend to draw more on the reserve in later periods because they are unable to keep the price from rising due to their limited ability to bank allowances.

For emitters, the consequences of non-compliance go beyond the financial penalties of acquiring additional allowances. Firms and other institutions receiving penalties face substantial costs of bad publicity as well. It is reasonable to expect that this prospect would cause many market participants to be quite risk-averse toward a compliance penalty. Having the PCR available as a last protection against non-compliance may allow risk averse participants to bid closer to their actual values for allowances due to their smaller risk of inadvertent non-compliance.

The next figure shows the levels of allowances sold or released from the reserve by period, but depending on whether the sale was post auction or in auction. This is the division tested in our price containment reserve treatments. The different pattern shows clearly the important difference one would expect to observe in the release of allowances from the California and RGGI reserves. The reserve design in these programs roughly corresponds to the difference between our treatments. Allowances are only released from the reserve during price spikes with the in auction reserve but are released in anticipation of future scarcity in the post auction sale. On average 18.75 permits were released in post auction sale sessions and 12.75 permits released for the in auction sessions. For the in

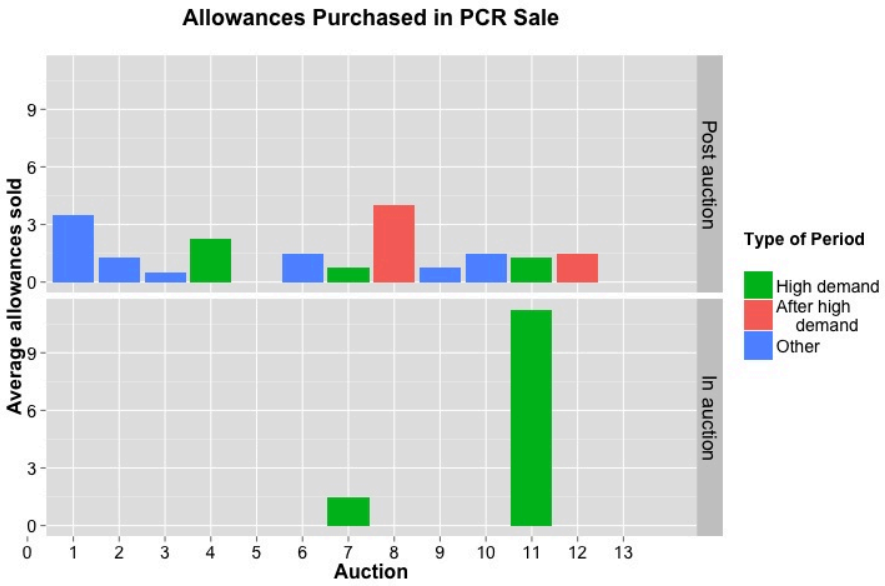


Figure 13: Permit purchases by period by type of PCR sale

auction sessions, all allowances released from the reserve were released during the last two of the three low hydro years, while for the post auction sale, the use of allowances is spread across periods.

It is not yet clear, whether the invariance of our market outcome measures to the type of release mechanism is due to some specific features of our experimental setup or may reflect some more general principle. We can see that different designs may have decidedly different implications for the pattern of release of allowances from the reserve, but if this does not necessarily imply differences in market performance, then there may be other reasons for preferring one type of mechanism over the other. The specifics of the implementation of mechanisms for releasing price containment reserve allowances into the market would appear to be a fruitful area for further research. That this distinctively different pattern of allowances being released from the reserve has so little apparent effect on liquidity, efficiency, or price is not unlike "the curious incident of the dog in the night-time."[28] It did nothing in the night-time and that was the curious incident. So it is here, the lack of noticeable effect of the reserve release mechanism merits a little additional sleuthing.

[28] From the short story, "Silver Blaze" in Arthur Conan Doyle, *Memoirs of Sherlock Holmes*.

References

Bodsky, R., Donato, D., James, K., & Porter, D. (2012). Experimental Evidence on the Properties of California's Cap and Trade Price Containment Reserve.: Economic Science Institute. Chapman University.

California Global Warming Solutions Act of 2006 (2006).

Dinan, Terry M. (2010). Managing Allowance Prices in a Cap-and-Trade Program. Washington, DC: Congressional Budget Office.

Fell, Harrison, Burtraw, Dallas, Morgenstern, Richard D., & Palmer, Karen. (2012). Soft and hard price collars in a cap-and-trade system: A comparative analysis. *Journal of Environmental Economics and Management, 64*(2), 183-198.

Goeree, J. K., Palmer, K., Holt, C. A., Shobe, W., & Burtraw, D. (2010). AN EXPERIMENTAL STUDY OF AUCTIONS VERSUS GRANDFATHERING TO ASSIGN POLLUTION PERMITS. *Journal of the European Economic Association, 8*(2-3), 514-525.

Hahn, R. W. (1984). Market Power and Transferable Property-Rights. *Quarterly Journal of Economics, 99*(4), 753-765.

Harris, Jeffrey H. (2010). Report on Holdings Limits To the Western Climate Initiative Markets Committee. In J. H. Harris (Ed.): Alfred Lerner College of Business and Economics. University of Delaware.

Hart, O.D. (1977). On the Profitability of Speculation. *Quarterly Journal of Economics, 91*(4), 579–597.

Hasegawa, Makoto, & Salant, Stephen. (2012). *Cap-and-Trade Programs under Delayed Compliance: Consequences of Interim Injection of Permits*. RFF Discussion Papers. Resources for the Future. Washington, DC.

Hothorn, Torsten, Hornik, Kurt , van de Wiel, Mark A. , & Zeileis, Achim (2008). Implementing a Class of Permutation Tests: The coin Package. *Journal of Statistical Software, 28*(8).

Jarrow, R. A. (1992). Market Manipulation, Bubbles, Corners, and Short Squeezes. *Journal of Financial and Quantitative Analysis, 27*(3), 311-336.

Monast, J., Anda, J., & Profeta, T. (2009). U.S. Carbon Market Design: Regulating Emission Allowances as Financial Instruments *Climate Change Policy Partnership* (pp. 36). Durham, NC: Duke University.

Montero, Juan-Pablo. (2009). Market Power in Pollution Permit Markets. *Energy Journal, 30*(Special issue 2), 1-28.

Pirrong, Craig. (2009). Market Oversight for Cap-and-Trade: Efficiently Regulating the Carbon Derivatives Market *Policy Briefs*: Brookings Institution.

Putniņš, Tālis J. (2012). Market Manipulation: A Survey. *Journal of Economic Surveys, 26*(5), 952-967.

Roberts, Marc J., & Spence, Michael. (1976). Effluent charges and licenses under uncertainty. *Journal of Public Economics, 5*(3-4), 193-208.

Stocking, A. (2012). Unintended consequences of price controls: An application to allowance markets. *Journal of Environmental Economics and Management, 63*(1), 120-136.

U.S. Energy Information Administration. (2013). State-Level Energy-Related Carbon Dioxide Emissions, 2000-2010. Retrieved 6/15/2013, from http://www.eia.gov/environment/emissions/state/analysis/

Van Horn, Andrew. (2013). *The PEAR/UVA Study: Testing AB 32's Market Design - Allowance Holding Limits, the APCR, CPUC Purchase Limits and the CA Allowance Strategy Calculator*. Van Horn Consulting. Orinda, CA. Retrieved from http://www.vhcenergy.com/sites/default/files/vhc_The_PEAR-UVA_Study-Holding_Limits-CA_Allowance_Strategy_Calculator-March_13_2013.pdf

Vossler, Christian A. (2013). Analyzing repeated-game economics experiments: robust standard errors for panel data with serial correlation. In J. A. List & M. K. Price (Eds.), *Handbook on Experimental Economics and the Environment* (pp. 89–112). Cheltenham, UK: Edward Elgar Publishing.

Weitzman, Martin L. (1974). Prices Vs. Quantities. *The Review of Economic Studies, 41*(4), 477-491