COLEMAN FUNG

RISK MANAGEMENT

RESEARCH CENTER

Working Paper # 2007 -08

Principal-agent incentives,excess caution, and market inefficiency: Evidence from utility regulation

Severin Borenstein, University of California Berkeley
Meghan Busse, University of California Berkeley
Ryan Kellogg, University of California Berkeley

December 2007

University of California Berkeley

Principal-agent incentives, excess caution, and market inefficiency: Evidence from utility regulation

Severin Borenstein, Meghan Busse and Ryan Kellogg

December 2007

Abstract: Regulators and firms often use incentive schemes to attract skillful agents and to induce them to put forth effort in pursuit of the principals' goals. Incentive schemes that reward skill and effort, however, may also punish agents for adverse outcomes beyond their control. As a result, such schemes may induce inefficient behavior, as agents try to avoid actions that might make it easier to directly associate a bad outcome with their decisions. In this paper, we study how such caution on the part of individual agents may lead to inefficient market outcomes, focusing on the context of natural gas procurement by regulated public utilities. We posit that a regulated natural gas distribution company may, due to regulatory incentives, engage in excessively cautious behavior by foregoing surplus-increasing gas trades that could be seen *ex post* as having caused supply curtailments to its customers. We derive testable implications of such behavior and show that the theory is supported empirically in ways that cannot be explained by conventional price risk aversion or other explanations. Furthermore, we demonstrate that the reduction in efficient trade caused by the regulatory mechanism is most severe during periods of relatively high demand and low supply, when the benefits of trade would be greatest.

Borenstein: University of California Energy Institute (www.ucei.org); Haas School of Business, University of California, Berkeley (faculty.haas.berkeley.edu/borenste); and National Bureau of Economic Research (www.nber.org). Email: borenste@haas.berkeley.edu. **Busse:** University of California Energy Institute and Haas School of Business, University of California, Berkeley. Email: meghan@haas.berkeley.edu. **Kellogg:** University of California Energy Institute and Department of Agricultural and Resource Economics, University of California, Berkeley. Email: kellogg@are.berkeley.edu. We are grateful for helpful contributions from Lucas Davis, Paul Gertler, Erin Mansur, Steve Tadelis, Matt White, and Frank Wolak and from seminar participants at U.C. Berkeley, U.C. Energy Institute, University of Michigan and NBER. We are grateful to the OpenLink Fund within U.C. Berkeley's Coleman Fung Risk Management Research Center for financial support.

1. Introduction

One of the central challenges that managers and regulators face arises when they must rely on agents, whose efforts and abilities are imperfectly observable, to choose actions that will advance the manager's or regulator's goals for the firm. The problem of designing incentive mechanisms that will induce economically efficient choices by agents is the subject of an entire field of economics. The usual solution when efforts and abilities are unobservable is to reward agents instead on the basis of observable outcomes. As Holmstrom (1999) discusses, one important question in this literature is how to get agents who are thus compensated to undertake the level of risk that is optimal from the principal's point of view. There is more than one possible impediment. One classic reason why agents may undertake insufficient risk is that they are simply more risk averse in income than the principal is. If this is the case, then an incentive scheme that rewards an agent according to the principal's payoffs will result in less risk-taking than the principal would prefer.

Holmstrom describes a second reason for an agent to undertake insufficient risk, one that does not hinge on risk-aversion over income: "career concerns." Career concerns arise when agents differ in their levels of ability, and when their rewards (compensation, retention, promotion) depend not only on the outcomes of their actions, but also on what the principal *infers* from those outcomes about the agents' underlying levels of ability. Examples in which rewards are based only on outcomes (and are therefore *not* influenced by career concerns) include sales commissions or any kind of piece-rate compensation: agents are paid on the basis of output, and the principal makes no attempt to assess how hard the agent worked or how skilled he or she is. Another example of a pure outcome-based reward is an exam in which a candidate receives a scholarship or gets admitted to a school based on his or her score. A third example is a fixed price construction contract in which a contractor is paid a previously-agreed sum once the building is satisfactorily completed, independent of when the job is finished.

In contrast, there are many examples in which career concerns do arise because rewards depend on *ex post* inference, rather than being determined solely by explicit functions of output. One example is professors receiving tenure. While more and better publications will help a professor receive tenure, there are generally not explicit contracts that specify how many papers of what quality will guarantee tenure. Instead, the tenure decision depends in large part on the senior faculty's inference about the candidate based on the candidate's output. Another example is executives who are promoted on the basis of job evaluations. Their quantifiable accomplishments matter, but so does what their boss thinks of them on the basis of their accomplishments. A third example is athletes drafted by sports teams: past performance will matter, but so will what the coach infers about the athlete as a result. A final example is startup firms funded by venture capitalists. The venture capitalists' continued investment depends not only on the financial returns, but on what they believe about the management team of the startup.

A key insight of Holmstrom's is that if agents recognize that their rewards depend in part on the principal's *ex post* inference, then they may try to manipulate the principal's formation of that inference. Speaking in terms of an agent delegated to make investment decisions, Holmstrom concludes that an agent will not prefer the investment opportunities that have the highest expected payoffs, but instead those that "leave him protected by exogenous reasons for failure." (Holmstrom, 1999, p. 179) The agent would like to choose a project for which, should it fail, the principal will have to take into account in forming *ex post* inferences some causes of failure that are not the agent's fault. Said another way, the agent "dislikes investments, which will reveal accurately whether he is a talented manager or not, since these investments make his income most risky." (Holmstrom, 1999, p. 179)

The idea that an agent may wish to undertake actions that will manipulate the *ex post* inferences made by an observer appears in a variety of contexts in the economics literature. For example, Brandenburger and Polak (1996) show that decision-makers whose rewards depend in part on what an outside observer ("the market") thinks of the decision will ignore their own contrary private information in order to choose what the market thinks is correct, a phenomenon they call "covering your posteriors." Scharfstein and Stein (1990) point to inference manipulation as a reason for managers to mimic the investment decisions of others —"follow the herd"—in order to signal to the labor market that they are "smart" managers who know what all the other smart managers know, rather than "dumb" managers whose information is just noise. Palley (1994) suggests that relative compensation will make it risky to make choices that are too different from the choices of others. In a model of "skill signalling," Harbaugh (2006) shows that an agent will avoid gambles for which loss has the potential to reveal poor judgment. A very different application of inference manipulation is Fudenberg and Tirole's (1986) "signal-jamming" model of predation, in which an incumbent firm tries to manipulate an entrant firm's *ex post* inference about the entrant firm's costs.

In this paper, we aim to extend this aspect of the principal-agent literature by examining not just the impact *on the firm* of decisions designed to manipulate a principal's *ex post* inference, but the impact *on the market* when many firms make decisions this way. We find that these effects can significantly distort market outcomes and create inefficiencies. In our particular context, this is exhibited as a reduction in the volume of transactions during precisely the times when potential gains from trade are highest: the market dries up just when trades would be most valuable.

The context in which we consider these issues is the market for wholesale natural gas. Over the last 30 years, regulators of public utilities have increasingly favored various forms of incentive regulation over simple cost-of-service regulation. This change in regulatory structure has brought to the fore the parallel between managerial and regulatory challenges in designing incentive-driven compensation. Regulators, who are the principals in this context, have two goals for the regulated utility. First, they want the utility to minimize its cost so as to minimize, ultimately, the costs to ratepayers. Second, they want the utility to ensure service quality. Of course, these two objectives will generally be in tension, because investments in service quality are likely to increase costs. The regulator generally gives the utility explicit incentives to reduce costs, spelled out in the rules governing the rate-making process. Incentives to provide service quality, however, may be more implicit. In this paper, we focus on the regulator's desire to have the utility avoid a very fundamental service failure: running so low on reserves that it has to curtail deliveries, temporarily denying customers access to natural gas. In the industry, this concern is referred to as "security of supply."

There are several reasons for regulators (and, as a consequence, utilities) to be concerned about security of supply in natural gas. First, there is very little scope for end-use customer demand response because regulated prices do not change very frequently, certainly not with day-to-day demand fluctuations. Second, spot markets for natural gas can be thin—especially during periods when demand is high—due to pipeline network constraints and coordination problems. As a result, if a utility has not procured enough gas to meet realized demand, it may not have a recourse other than curtailment to address the shortage in the short run. A curtailment is something regulators—and utilities—want to avoid, both because of economic costs (such as business disruption for industrial customers) and political costs (arising from inconvenienced customers complaining to elected political leaders who then demand accountability from regulatory bodies).

As we have spoken to utility executives, it is clear that they believe that there would be significant negative consequences for this type of service failure. Furthermore, it is also clear that utilities believe that the harshness of the penalty depends on what inference the regulator draws about *why* a curtailment happened. In particular, a curtailment that occurred despite a utility's best efforts to procure additional supplies would incur consequences less severe than a curtailment that was more directly attributable to a mistake or misjudgment on the utility's part. In light of this, utility executives have told us that they would particularly wish to avoid a situation in which they had sold reserves of gas that could have been used later to avert or mitigate a curtailment. Not only would these circumstances appear to the ratepaying public as (at best) a lack of foresight or (at worst) profiteering, but it virtually guarantees that the regulator will not infer that the utility was doing its best to avert the disaster.

We predict that utilities, in an attempt to prevent the regulator from having grounds to make this inference, will resist selling reserves when it expects the spot market to be tight and curtailments to be possible, and that it will do so to a greater degree than the regulator would wish it to. We call this behavior "excess caution." Of course, the regulator wants the utility to be cautious about making sales in such a situation, because curtailments are costly for the regulator. By "excess caution" we mean the additional reluctance on the utility's part to sell reserves it might need, in order to manipulate the regulator's inference process.[1] Excess caution in this context is analogous to the unwillingness to invest on the part of Holmstrom's (1999) managers, the desire to herd into popular investments by Scharfstein and Stein's (1990) managers, or the choice to do what "the market" wants by Brandenburger and Polak's (1996) decision-makers.

Our primary aim in this paper is to model and empirically estimate the effect of security of supply concerns and excess caution on market outcomes in wholesale natural gas markets. Our model yields two testable market implications. First, the forward price for natural gas will exceed the expected spot price when supplies are expected to be tight and curtailments are possible. Second, because utilities will be relatively less likely to sell gas in forward markets in tight-supply conditions, even when there exist other firms that place a higher valuation on marginal supplies, the quantity of forward transactions will be reduced. We find that data from natural gas markets support both of these predictions. Furthermore, while alternative theories regarding firm behavior in these markets, such as price risk aversion, can also predict the existence of forward premia, only "excess caution" of utilities regarding forward sales of gas implies reductions in forward transaction quantities.

Our findings suggest that skill signalling between a principal and agent, in combination with the agent's expectation of a severe penalty for being judged as relatively unskilled, can cause significant market inefficiencies beyond the individual principal-agent relationship. In natural gas markets, it appears that the desire of utilities to avoid curtailments that

[1] The caution that regulators would wish to have exercised would result, on the margin, in an equal reluctance to sell and eagerness to buy; i.e., equal marginal valuations for an incremental sale as for an incremental purchase. "Excess caution" is a greater marginal valuation for incremental sales than for incremental purchases; i.e., a greater reluctance to sell than willingness to buy. In section 3, we incorporate this concept in a theoretical model.

could be directly attributed to their actions actually makes it substantially more difficult for utilities as a group to allocate supplies to those utilities and customers who need them most. Furthermore, this reduction in efficient trade occurs at precisely those times at which the trades are most valuable: when markets are tight and the threat of curtailment is most severe.[2]

"Excess caution" is not unique to regulated industries, or to natural gas markets. In a business context, such behavior is sometimes referred to colloquially as CYA ("cover your ass") behavior. During the 1980s, when IBM dominated the emerging personal computer market, this behavior was embodied in a phrase that was known by every corporate purchasing agent: "No one ever got fired for buying IBM." Taking the risk of purchasing a different brand of PC was seen as having tangible downside for the purchasing agent if the alternative machine performed poorly, and little upside if it resulted in overall cost savings.

In what follows, we first describe in section 2 the relevant institutional details of the natural gas industry. In section 3 we present a theoretical model of forward and spot wholesale gas procurement that draws from the institutional setting. Section 4 discusses empirical tests of the model's prediction that forward price premia will arise in markets that are expected to be tight, and section 5 tests the prediction that forward trading volumes will decrease in tight markets. Section 6 presents some additional, supplementary evidence, while section 7 concludes and discusses broader implications.

2. The Natural Gas Industry

2.1. Market Structure

The wholesale natural gas industry can be broadly classified into three main segments: production, transportation, and local distribution. Natural gas production, which is generally considered to be competitive,[3] occurs mainly in a belt running northwest to southeast from the Rocky Mountains to the Gulf of Mexico. The demand centers for natural gas, however, are located primarily in the Northeast, Upper Midwest, and West Coast. Thus,

---

[2] Lyon (1992) models "hindsight review," the practice of disallowing utilities from recovering fuel costs if the procurement contract chosen by the utility (forward vs. spot) is *ex post* not the least expensive choice. While both models are concerned with choices about spot vs. forward procurement, and both involve an *ex post* evaulation by the regulator, the two papers are otherwise about different aspects of the procurement and regulatory review processes.

[3] Wellhead natural gas price deregulation began in 1978 with the Natural Gas Policy Act. Prices were fully de-controlled in 1993 under the Natural Gas Wellhead Decontrol Act.

a network of interstate gas transmission pipelines has been developed to transport gas to where it is needed. Each pipeline is a distinct private legal entity, though some entities are jointly owned by a single holding company.

Interstate natural gas pipelines are regulated by the Federal Energy Regulatory Commission (FERC). According to FERC regulation since 1992, pipelines are not permitted to buy and sell natural gas; they may only act as the transporter of gas on behalf of their customers, known as shippers, who hold title to the gas itself. Shippers may be upstream gas producers, downstream gas consumers, or gas marketers.

Via these transportation arrangements, gas is delivered to wholesale buyers, which are primarily regulated local distribution companies (LDCs) responsible for gas distribution to ratepaying customers.[4] The LDCs are regulated by state Public Utilities Commissions (PUCs) with two primary objectives: (1) to ensure reliable gas supply so that customers will not be interrupted, even during peak demand periods, and (2) to minimize customer rates while allowing the utility to achieve a reasonable rate of return. The prices LDCs charge to their customers are regulated through traditional cost-of-service regulation, under which the wholesale cost of gas supply is passed through to ratepayers, though lags in regulatory adjustment of retail prices and review of abnormal prices give LDCs incentives to reduce their natural gas purchase costs.

2.2. An LDC's Gas Procurement Decision

Given the natural gas market structure, an LDC has numerous gas procurement options at its disposal. Broadly speaking, it must make its decision along two dimensions: (1) whether to take ownership of wholesale gas "downstream" at the location of its customers or at an "upstream" producing location, and (2) how far in advance to procure. We now consider the institutional aspects of these decisions in turn.

Should the LDC elect to procure gas downstream, its procurement process is relatively simple—it must contract with a gas supplier for delivery in its local area. In this case, it does not need to contract for transportation capacity on an interstate pipeline: it is the responsibility of the gas supplier to ensure transportation to the point of sale. Alternatively, if the LDC arranges for gas delivery in an upstream producing zone, it is then also responsible for contracting sufficient transportation to ship the gas to its local area. In either case, coordination is necessary amongst the three market participants—the LDC, the gas seller, and the pipeline—to ensure that the appropriate quantities are delivered

[4] Some merchant electric generators and large industrial firms also purchase gas directly in wholesale markets.

at the appropriate times. The difficulty of achieving this coordination, which includes the search costs of finding a suitable counterparty for each transaction, may be significant, particularly in high demand periods when pipelines are operating at their capacity constraints.[5]

LDCs also have several options for the timing of their purchases. Day-ahead spot markets for gas currently operate in approximately 100 locations across the U.S., and are liquid to varying degrees: while the day-ahead market at Henry Hub in Louisiana (the delivery point for gas traded under a NYMEX monthly natural gas futures contract) is regarded as deeply liquid, industry participants have told us that at many locations, firms may sometimes be unable to purchase gas on the spot market, particularly in tight market periods, because liquidity is limited by pipeline capacity constraints and coordination problems. Thus, in a high demand environment in which spot markets are expected to be tight—for example, due to cold weather in the Northeast—an LDC faces a risk that it will be unable to procure gas when it needs it.

As an alternative to participating in day-ahead spot markets, an LDC can also procure gas in a forward market called the "bidweek" market. "Bidweek" refers to the last five trading days of each month. Transactions in this market are for natural gas delivery in local gas markets for the coming month. Bidweek transactions are for a specified volume of gas every day over the month. An LDC could make a bidweek purchase of gas delivered to its local market, or could combine a bidweek purchase of gas at an upstream location with transportation rights on a pipeline that can carry the gas to its local market.[6] By purchasing gas in these forward markets, an LDC may avoid reliance on the uncertain spot markets for its gas supply. LDC's can also arrange for gas and for transportion with long-term contracts. In this paper, however, we will be concered with the spot and forward markets (namely, the bidweek and capacity release markets) described above.

3. A Model of Security of Supply and Excess Caution in Gas Markets

In this section, we model the decisions of natural gas local distribution companies (LDCs) to purchase natural gas supplies in forward and spot markets. We show that natural gas

[5] Apart from the NYMEX natural gas futures market, local natural gas and pipeline capacity markets are bilateral trading systems without a central market-maker.

[6] An LDC may own pipeline transportation rights under a long-term contract, or it can acquire temporary rights as a "capacity release" from another transportation rights holder. The standard terms for a capacity release are for the shipment of a specified volume of gas every day for a calendar month, terms that parallel the bidweek contract for gas. The capacity relase market is considered in more detail in a later section.

markets can exhibit both a forward price that exceeds the expected spot price and reduced transaction volumes when spot markets are expected to be tight, if two conditions are met. The first condition is that LDCs exhibit "security of supply" concerns, meaning that having insufficient supplies of gas is more costly than having comparable excess supplies because the former implies a risk of curtailment. The second condition is that LDCs behave with "excess caution" in their forward sales, meaning that they do not want to have sold gas they later needed. Furthermore, we demonstrate that, absent "excess caution," a security of supply model alone could be consistent with the presence of forward price premia in tight markets, but that it alone is unlikely to explain reductions in forward transaction quantities. Finally, we examine a model of price risk aversion, which also does not predict a reduction in forward transactions in tight markets, and which, furthermore, seems inconsistent with the institutional structure of the natural gas industry.

3.1 Supply and Demand Model With Security of Supply

We begin by constructing a model of supply and demand in natural gas markets in which LDC's have security of supply concerns because they face asymmetric costs of having too much vs. too little gas, but in which they do not behave with "excess caution." Then, in section 3.2, we add the effect of "excess caution," namely, a reluctance to sell reserves of gas that arises from the utility's desire to keep the regulator from being able to blame it for a curtailment. (An example of a principal-agent mechanism showing how excess caution arises is presented in appendix 2.)

Consider a model in which several wholesale gas buyers—LDCs who need to meet end-use customer demand in the coming month—are present at the downstream end of a pipeline (a.k.a. the "local market"). As a result of either past purchases or previously established contracts, the LDCs will hold some "reserves" of gas that can be used to meet this upcoming demand. "Reserves" may include gas in local storage that the firm already owns or has purchase contracts to obtain in the coming month, and also gas located upstream (that the firm owns or has contracts to obtain) paired with the necessary pipeline transportation rights to move the gas to the local market. If the current level of reserves does not equal the level of supplies that the firm would ultimately like to have on hand in the coming month given the anticipated customer demand, then the firm can use the forward or spot markets to buy additional gas or sell its excess supply.

In the model, customer demand varies exogenously due to weather and other shocks, and is extremely inelastic in the short run. This closely mirrors reality in natural gas markets, where prices for end-use customers are determined by regulators, and change infrequently and with a lag. As a result, LDCs cannot rely on end-use price response to reconcile customer demand with their supplies.

We assume for now that LDCs are risk neutral with respect to price—an issue to which we return in section 3.3—and model LDCs' demand for wholesale gas given that they face asymmetric costs of mismatching their supplies of gas to realized customer demand. Specifically, shortfalls, which result in curtailment of end-use customers, are much more costly than excess supplies, which are generally storable at some location. This asymmetry leads an LDC's demand curve for supplies of gas during any given month to be convex: the marginal value of supplies is very high for units up to the realized customer demand that the LDC faces, and flattens out rapidly for units beyond that level. Figure 1 shows hypothetical marginal valuation curves for supplies of gas in a particular delivery month once the demand shock has been realized (and final customer demand is known).[7]

If an LDC could always be certain that it could adjust its level of supplies to exactly match its customers' demand through use of the spot market, this asymmetry would be of no consequence.[8] However, we have been told by many market participants that the — market for natural gas is sometimes not sufficiently liquid for LDCs to be certain that they can buy or sell the quantities they wish to trade. For this reason, LDCs procure most of their supplies in the forward markets, usually between one week and six months in advance.

Forward prices if no spot market were available. If LDCs turn to the forward market because of concerns about an illiquid spot market, how does that affect the forward market equilibrium? In order to highlight the role of spot market access in the answer to this question, we first consider the extreme case in which an LDC has no access to a spot market, meaning that no adjustments are feasible after demand shocks are revealed. In this case, an LDC must buy whatever gas it needs in the forward market, before the retail state of demand is completely known. When facing the prospect of buying (or selling) in the forward market, the LDC's marginal valuation for each unit of gas will be the weighted average of the marginal valuations it would have for that unit of gas under

[7] Though the effect appears to be similar, the convexity of the marginal valuation curve here is not due to risk aversion. The firm and its decisionmakers are assumed to exhibit a constant marginal valuation of wealth. Rather, the asymmetry in monetary payoffs from excess versus insufficient supplies yields the convex marginal valuation curve.

[8] For example, consider the case of a long-distance trucking firm. If the firm were to have insufficient fuel, it would cripple the firm's ability to perform. However, long-distance trucking firms can rely on a dense, geographically diverse, and very reliable spot market for fuel in the form of gas stations and truck stops. They do not need to make forward purchases of fuel in order to avoid security of supply risk. Thus, if the input spot market is sufficiently thick, we expect that it would be a good substitute for the forward market, even for firms that have security of supply concerns.

all possible retail demand shock realizations. Figure 2 illustrates a hypothetical forward marginal valuation curve, $\bar{D}(q)$, if there were two equally likely retail demand realizations, $D(q, w_1)$ and $D(q, w_2)$ where w_1 and w_2 represent high and low demand shocks (due, for example, to cold and warm weather). We use the term marginal valuation curve, rather than demand curve, to emphasize the fact that the LDC may either buy or sell in the forward market, depending on the level of reserves it controls coming into this market. This point is illustrated in figure 3. Suppose that a hypothetical firm has q_{res} in reserves as it comes into the forward market. If the price in the forward market is above the firm's marginal valuation at its current level of reserves, q_{res}, the firm will want to sell reserves according to the $\bar{D}(q)$ schedule. If the forward price is below this level, the firm will want to buy. That is, the firm's $\bar{D}(q)$ becomes part of the market supply curve for prices above $\bar{D}(q_{res})$ and part of the market demand curve for prices below $\bar{D}(q_{res})$.

Figure 4 illustrates hypothetical supply and demand curves for firms that are holding different levels of reserves relative to their expected needs. Panel (a) shows supply and demand curves for two firms with abundant reserves relative to the expected demand realization. These are firms that would be likely to have excess supply for the current period if they were to continue with their current level of reserves; the supply and demand curves are therefore relatively flat, up to the point where the firm is considering selling a large amount of reserves. Panel (b) shows two firms with low levels of reserves relative to expected demand. These are firms that would be more likely to experience shortage if they were to continue with their current level of reserves. These firms have steeper demand and especially supply curves, reflecting a much higher possibility of curtailment than for the firms in panel (a). (The market supply and demand curves in the forward market are obtained by aggregating the individual supply and demand curves, such as those depicted in figure 4, across the firms in the market.)

How will an LDC's purchases or sales of gas on the forward market in this case differ from what it would choose in the spot market, were the spot market available to it? In appendix 1, we show that convexity of the marginal valuation curves—in conjunction with another plausible, though not general, condition on the functions—ensures that the quantity that the LDC would choose to acquire in the forward market, through purchase or sale, at a given forward price p will be *larger* than the expected quantity it would choose to purchase in the spot market (were the spot market available to it) at the same price p. Or, equivalently, if $E(q)$ is the LDC's expected spot market demand at spot price p, the LDC's marginal value of reserves in the forward market at quantity $E(q)$, namely $\bar{D}(E(q))$, will be greater than p. Figure 5 illustrates these relationships for a hypothetical example. The intuition of this result is that the asymmetric costs of shortage versus excess supply—reflected in the convexity of the marginal value curve—gives strong incentives to purchase more than will be needed *on average* in order to minimize the expected costs of supply/demand mismatches. As discussed in appendix 1, when the marginal valuations of all LDCs in the market share this property, in equilibrium the forward price can exceed the expected price that would occur in the spot market, were the spot market available.

Forward prices if spot market may be available. What happens if, contrary to what we have assumed so far, the LDCs *do* have access to a spot market? If the spot market is sufficiently liquid that an LDC can always rely on spot market transactions to supply any additional gas it needs during the delivery month (or provide buyers for any excess gas), then no LDC would ever purchase in the forward market at a premium over (or sell at a discount to) the expected spot price, so the forward price, p_f, would equal the expected spot price, $\bar{p}$. Suppose, however, that there is some chance that an LDC will not be able to utilize the spot market, perhaps because the spot market is thin and the LDC may not be able to find a counterparty with whom to trade within the necessary time frame. If the LDC is unable to buy in the spot market, the loss associated with the purchasing failure is equal to the expected marginal valuation of the gas that the LDC did not get to buy. Thus, in the forward market, the LDCs would be willing to pay no more than the weighted average of the expected spot price (what it will pay if it *can* access the spot market) and the expected marginal valuation (what it will lose if it *cannot* access the spot market), where the weights are the probabilities of being able to access the spot market or not. This is illustrated in figure 6. If the demand conditions discussed above (and in the appendix) are present, then the willingness to pay in the forward market for the expected quantity the firm would purchase at a particular spot price will still exceed that spot price. In this way, security of supply concerns can lead to a premium in the forward price over the expected spot price.

The model we have presented so far indicates not only that a forward premium can exist, but also that its size is likely to vary with the expected tightness of the gas market. If all or most market participants have plenty of reserves on hand in anticipation of upcoming needs, then the market supply and demand curves will be relatively flat at a level very close to the expected spot price, $\bar{p}$. (Furthermore, $\bar{p}$ will also be very close to the carry-forward value of gas; namely, the expected spot price next period minus marginal storage cost). If, however, the spot market is expected to be tight, then many participants' q_{res} will be at a point where their marginal value is far enough above the expected spot price that—considering the risk of being unable to transact in the spot market—they are willing to pay (or only willing to accept) a price that is significantly above the expected spot price. These two different situations are illustrated in figure 7.[9] This is not a proof that the forward premium would necessarily increase with market tightness, but the theory suggests that is a likely outcome.

Forward quantities. As explained earlier, we are interested in studying both prices and transaction volumes in the forward market. The model presented so far also helps to analyze the impact that security of supply concerns will have on volumes. If each market participant has a forward market marginal value function $\hat{D}(q)$ and an initial endowment of gas q_{res}, then if the equilibrium forward price is p_f, each participant will want to buy or sell a quantity Δq in the forward market until the marginal value at its resulting level of reserves is equal to the market price, $\hat{D}(q_{res} + \Delta q) = p_f$. From a market-wide perspective, the trading volume will be that which equalizes the marginal valuations of all market participants; *i.e.,* , volume is equal to the mismatch between the initial reserve allocation and the efficient equilibrium allocation after the forward market has cleared.[10]

For three reasons, which we discuss in more detail in the appendix, we expect that the mismatch between firms' initial and equilibrium allocations will be greater, and the trading volume therefore also greater, at times when the spot market is expected to be tight: (a) tight markets are associated with a larger overall gas market due to high demand, which will tend to scale up the size of the mismatch (scale effect); (b) tight markets are associated with weather shocks that impact neighboring areas heterogeneously and change the need for reserves unevenly because some firms now need much more gas, but others do not (heterogeneity effect); and (c) any transaction cost of trading is more likely to be overcome at peak times because high demand shocks will shift marginal value functions to the right, increasing the likelihood that the current level of reserves a firm holds lies

[9] The same argument suggests that while the asymmetric costs of insufficient versus excess supplies tends to create a forward premium when markets are tight due to concerns about security of supply, it does not create a forward *discount* when markets are loose due to concerns about "security of demand." It is indeed the case that in loose spot markets, a seller of reserves might be unable to find a buyer, but the loss from failing to complete a transaction will be minimal. While an LDC that is short of reserves may be desperate to procure supplies in order to prevent customer curtailments, an LDC that has excess reserves will not be desperate to sell. Its downside risk is limited to foregoing potential revenue from selling the reserves at the market price. In most cases, it could simply carry the reserves forward to the next period. As a result, the supply curve is relatively elastic (as illustrated in panel (a) of figure 4), implying small surplus gains from making a sale. In addition, because sellers are also mostly regulated LDCs who pass their net gas procurement costs through to end-use customers, the financial risk for the LDC itself is limited even further.

[10] Stated more precisely, the trading volume is the sum of the differences between the initial allocation and the efficient equilibrium allocation over all firms whose initial allocations exceed their equilibrium allocation. This will also equal the sum of the differences between the equilibrium allocation and the initial allocation over all firms whose initial allocation is less than their equilibrium allocation. In other words, total sales are equal to total purchases are equal to the total volume of trade.

in the steep portion of the marginal valuation curve, thereby making greater marginal value differences between buyers and sellers more likely (transaction cost effect). While we don't know the magnitudes of these effects, they all imply greater forward market volume when demand is expected to be high and spot markets are therefore expected to be tight. Together, they suggest that the security of supply model alone would not be consistent with a finding of decreased volume during tight-market periods.[11]

3.2 "Excess Caution" Induced By A Principal-Agent Mechanism

While the model of "security of supply" does offer insights into the effect of asymmetric costs of shortages versus surplus supplies, it does not yet capture one of the institutional aspects of gas utility regulation that has been expressed or confirmed to us by all of the participants we have interviewed: regulatory oversight in the retail natural gas market results in especially harsh treatment of an LDC that *sells* supplies it had in reserve and then finds itself in a shortage situation and in need of those supplies. The underlying reason for such a penalty in this industry is most likely political. The regulator is ultimately accountable to the governor and legislature, and they in turn are accountable to their constituents. While curtailments will attract customer ire whenever they occur (hence the convexity of the marginal valuation curve), the political fallout will be much greater if a utility had necessary reserves and then sold them, which *ex post* looks like "profiteering," than if the utility can argue that it made a "good faith effort" to acquire additional supplies, but that the market was simply too thin or the coordination problems too difficult to surmount in a short period of time. Where the political fallout is greatest, the political pressure on regulators will be greatest, and so, therefore, will the regulator's punishment of the utility be harshest. Although these punishments are not easily specified, nor are they written in formal rules, it is clear that the utilities believe they are real. One utility executive told us with regards to this issue, "[Avoiding] regret is the prime mover" in dealing with regulators.[12] The resulting incentive is for utilities to exercise "excess caution" by holding onto reserves rather than selling them, so as not to be demonstrably blameworthy for a shortfall.

---

[11] These arguments focus on buyers and sellers of physical gas, as opposed to price speculators who plan to unwind their positions in the spot market instead of completing the forward agreement. To the extent that speculators are present in these markets, it seems likely that they too would be more active in tight-market periods when there is more price volatility and the value of superior market information is greater.

[12] Although the circumstances are quite different, the blackouts during the California electricity crisis and the subsequent recall of Governor Gray Davis suggest that the political consequences of perceptions of poor oversight of public utilities can indeed have substantial political consequences.

We can represent this feature of the market graphically with a penalty function that raises the reservation price at which an LDC would sell reserves in the forward market if there is a risk it will then have to procure supplies in the spot market in order to meet its own demand. The expected penalty, and associated increase in the reservation price in the forward market, is greater if the LDC is more likely to have to buy quantities back in the spot market. This likelihood is a decreasing function of its starting reserve position and an increasing function of the quantity of gas it sells in the forward market. We illustrate two scenarios in figure 8, where in each case $\hat{D}(q)$ represents the willingness to buy and sell in the forward market absent the regulatory penalty, and $\tilde{D}(q)$ is the willingness to buy and sell in the forward market inclusive of the regulatory penalty. In figure 8a, the LDC has plentiful supplies and could sell substantial quantities in the forward market before the expected value of the penalty would become non-negligible. In figure 8b, however, the LDC is in a tighter reserve position and is more likely to need to purchase supplies in the spot market. In this case, selling even a small quantity in the forward market creates a significant probability that the LDC will incur regulatory punishment. Note that this creates a discontinuity in the marginal value of reserves at the LDC's starting reserve position.

The regulatory penalty causes firms' supply curves, such as those illustrated in figure 4, to shift up while having no effect on their demand curves. The result is a larger forward premium, and the effect is much more likely to occur when supplies are expected to be tight. In fact, in a loose market with plentiful supplies, the penalty threat will likely have no effect.

A second result of the penalty threat is to lower volume in the forward market when the threat of the regulatory penalty is most salient. We argued in the previous subsection that, absent such a regulatory penalty, an expectation of tight spot markets implies a forward trading volume as large as or larger than the forward trading volume when spot markets are expected to be loose. However, the asymmetry of the regulatory penalty—among any LDCs who end up with insufficient supplies in the spot market, it will affect only those who sold gas in the forward market—acts to discourage efficient forward trade when spot markets are expected to be tight. A sufficiently large penalty could eliminate all trade in the forward market at times when LDCs perceive a real risk of a severe demand shock and tight gas market.[13]

---

[13] Since reduced volume and higher prices are often a hallmark of market power, and since a tight forward market also suggests a reduced number of market participants, one might wonder whether market power in local markets might be an alternative explanation. Market power might exacerbate the effect if a reduced number of participants in the forward market gives firms the opportunity to exercise (or more

One might wonder why the regulator would impose a penalty threat that causes the LDC to forego efficient trades in the forward market and instead exercise "excess caution." In appendix 2, we show that, in a principal-agent framework in which the principal is uncertain of the agent's ability to trade efficiently on its behalf, an optimizing principal may want to impose a penalty on *ex post* unfavorable outcomes of an action, but not of an inaction. The principal may prefer to do so even though the policy would in equilibrium lead to "excess caution." This is not to say that the regulatory penalty that LDCs claim to face in the natural gas industry is necessarily the result of optimizing behavior, but that it is also not necessarily irrational.

In the discussion of security of supply and excess caution we have focused on the behavior of local distribution companies that purchase for resale to residential, commericial and industrial end users. Some large industrial gas consumers, however, do not buy through LDCs, but instead participate in the wholesale gas market directly. This includes electricity generating companies as well as energy-intensive manufacturers, such as cement producers. Because gas is a critical input to their production processes, these companies are also likely to have convex marginal value functions. The regulatory penalty discussion would not apply directly, but equivalent principal/agent conflict inside the firm could yield excessively cautious behavior. A manager responsible for gas procurement could face very similar incentives not to resell reserves, even at a high incremental profit to the firm, if there is even a small probability that the firm will end up needing the gas and be unable to procure it on the spot market.

3.3. Price Risk Aversion

Thus far, we have assumed that the market participants are risk neutral with respect to price. However, price risk aversion could also affect the relationship between the spot and forward markets. Buyers and sellers who are risk averse will be willing to give up some expected surplus in a transaction in order to reduce the price volatility they face. If buyers are systematically more willing to pay to reduce price uncertainty than are sellers, then the observed forward transaction prices can be greater on average than the associated spot prices.[14]

Risk aversion over prices, however, is not as compelling an explanation for a forward

---

effectively exercise) market power by witholding quantity to drive up price. However, there has to be some reason for fewer LDC's to be willing to participate in a forward market when spot prices are expected to be high. We don't see a route whereby market power alone would have this effect.

[14] See Dusak (1973), Bodie and Rosansky (1980) and Carter, Rausser and Schmitz (1983) for a more thorough discussion of risk premia in efficient and liquid commodity markets.

price premium in these markets as it might at first appear. First, recall that buyers and sellers in downstream gas markets are generally the same firms. A holder of a certain quantity of long-term gas delivery contracts will sometimes find itself in need of more, and thus buy on the local gas market, and at other times find itself with excess supply and therefore sell on these markets. It is difficult to argue that buyers are systematically more averse to price volatility risk than sellers when the same set of players occupy both sides of the market at various times. In addition, the LDCs that are buying and selling in this market are generally allowed to pass through nearly all of their gas acquisition costs to their customers, so we would not expect them to exhibit much price risk aversion.

While these institutional factors reduce the likelihood that risk aversion is a factor in these markets, we still assess its implications for the relationship between forward and spot prices, and in the quantity of trades. A simple model demonstrates how price risk aversion could affect the market. Assuming that there is a common distribution of beliefs about the probability distribution of the spot price, with an expected spot price of $\bar{p}$, risk-averse buyers will be willing to pay more than $\bar{p}$ to hedge their purchase costs in the forward market and risk-averse sellers will be will to receive less than $\bar{p}$ in order to hedge their sales revenues in the forward market. Risk-neutral participants stand ready in the forward market to buy at any price below $\bar{p}$ and sell at any price above $\bar{p}$. Finally, risk-averse speculative buyers would only buy in the forward market if they could do so at a price below $\bar{p}$ and risk-averse speculative sellers would only sell in the forward market if they could do so at a price above $\bar{p}$.[15]

A hypothetical forward market with large numbers of risk-neutral buyers and sellers is illustrated in figure 9 by the D_f and S_f curves. To the left, downward sloping demand at prices above $\bar{p}$ represents differences in risk aversion among hedging buyers. To the right, downward sloping demand at prices below $\bar{p}$ represents differences in risk aversion among speculative buyers. On the supply curve, risk-averse hedging sellers are on the left and risk-averse speculative sellers are on the right. We assume there is a very small positive trading charge so that risk-neutral buyers and sellers do not trade with one another. With a large number of risk-neutral traders, supply and demand will intersect at a price near $\bar{p}$, with the marginal trade very likely occurring between a risk-neutral seller and a hedging buyer, or between a risk-neutral buyer and hedging seller. (Figure 9 depicts the latter

[15] We carry out this analysis considering the variance of spot price as the relevant measure of risk. If the participants could actively engage in diversification of this risk, then the CAPM would apply and only the non-diversifiable risk would be relevant. Given that we are considering local gas markets, this seems less plausible. Still, the analysis goes through unchanged with substitution of non-diversifiable risk for total risk.

case.) In equilibrium, hedgers on both sides of the market are able to lock in a price in advance while sacrificing essentially no expected surplus. Risk-averse speculators do not find any attractive trading opportunities and do not trade in the market.

Note that for a given $\bar{p}$, if the variance of the expected spot price distribution changes, the willingness-to-pay or willingness-to-accept of the risk averse market participants will change, but not of the risk-neutral players. An increase in spot price variance causes the risk premia that market participants are willing to pay or risk discounts that they are willing to accept to increase. So, the parts of the supply and demand curves that are not flat become even steeper, as illustrated by the dashed lines in figure 9, D'_f and S'_f. The volume of trade will be essentially unchanged, and there still will be no significant forward premium or discount.

Now consider the forward market in the absence of risk-neutral traders. The flat areas of supply and demand that were in figure 9 disappear. In figure 10, the demand curve above $\bar{p}$ is made up of hedging buyers and below $\bar{p}$ consists of speculative buyers. The reverse is true of the supply curve; hedging sellers are below $\bar{p}$ and speculative sellers are above $\bar{p}$. Figure 10 shows a situation in which the equilibrium price in the forward market happens to equal the expected spot price. In this special case, an increase in spot price volatility leaves the forward premium unchanged at zero. Both the supply and demand curves get steeper, but they rotate around $\bar{p}$ so the equilibrium price and trade volume are unchanged.

Figure 11 illustrates the situation in which supply and demand for forward trades from risk-averse participants, combined with the absence of risk-neutral participants, results in a forward premium.[16] On the margin, speculators on the sell side of the forward market are transacting with hedgers on the buy side. In such a situation, an increase in the volatility of the spot market will rotate the supply and demand curves around the $\bar{p}$ line and will increase the forward premium.[17] Thus, an increase in spot price volatility could increase a forward premium that is already present in the market.

The price risk aversion model, however, has a different prediction for the forward trade volume than our model of security of supply with excess caution. In a simple mean-variance utility model, the increase in spot price volatility would rotate the forward demand and

---

[16] This outcome could also occur if there were risk-neutral traders as long as they in aggregate were not willing to supply or demand a large quantity relative to the market.

[17] And conversely, if the forward market displays a discount to the expected spot price, an increase in the spot market volatility will increase the size of the forward discount.

supply curves upward (for prices above $\bar{p}$) by exactly the same proportion. This follows immediately from the assumption that the disutility of spot price variance changes linearly with the variance. In this case, the change in spot price volatility has no impact on forward market volume. The same two traders as before would be on the margin. All traders who were in the market at the lower volatility level would still be in when the expected volatility of the spot market increases, and all those who were out would still be out.

In a more complex model of the impact of risk on utility—such as a model in which utility is a higher-order polynomial function of spot price variance—this would not necessarily be the case. However, deviations from the volume-neutrality of changes in spot price volatility would depend on higher-order terms in the utility model causing the supply and demand curves to rotate upward by different proportions around the equilibrium. The sign of such a change in volume would be ambiguous. Even in that case, however, the volume would almost certainly be bounded between the quantities at which the supply and demand curves cross $\bar{p}$.[18]

While the theoretical model of price risk aversion narrows substantially the range in which forward market volume may vary with the expected tightness of the spot gas market, two factors not explicitly modeled suggest that volume is likely to increase when the spot market is expected to be tight. These two factors are similar to those that affect volume in the security of supply model. The first is again a scale effect: if the gas market is larger at times when demand is high and the spot market is expected to be tight, the demand and supply of hedging or speculating traders in the forward market is likely to scale up with market size, increasing the total volume of trade. The second factor is a possible need to overcome transaction costs. As illustrated in figures 9, 10 and 11, increases in volatility that are associated with a tight gas market increase the slope of supply and demand in the forward market. Transaction costs act like a tax, creating a strictly positive minimum surplus gain necessary to justify a transaction. Increases in volatility will therefore cause the gains from trade to exceed the transaction costs more frequently, driving an increase in trading volume.

Overall, price risk aversion—the most common explanation for forward-spot expected price differences in finance—seems unlikely to explain a significant forward premium in this instance for institutional reasons. Still, if forward premia in the natural gas industry are caused by price risk aversion, theory suggests that the premia could very well increase with expected volatility in the spot market: a condition that would be expected to occur

[18] This requires that the slopes of the supply and demand curve do not change signs and that changes in spot price *volatility* do not induce a hedger to become a speculator or vice versa.

when supplies are expected to be tight. However, the same theory would be hard-pressed to explain a significant decline in forward transaction *volumes* at such times.

3.4. Theory Summary

We have presented three models of forward market transactions in the wholesale natural gas market. All three models—security of supply concerns, regulatory/organizational incentives for excess caution, and price risk aversion—can predict a forward market price premium that increases with the expected tightness of the spot market for gas. For institutional reasons, however, the price risk aversion model seems relatively unlikely to apply in this setting.

The implications of the models for transaction volumes differentiate them more distinctly. Security of supply concerns, absent excess caution, suggest that forward volume will increase when the spot market is expected to be tight. Forward volumes are also likely to increase in such markets if price risk aversion is important. Though it is difficult to infer under either model the degree to which volume would be expected to increase, neither seems consistent with a significant drop in forward market volumes when spot markets are expected to be tight. This inconsistency contrasts with the prediction of the excess caution model. The excess caution theory predicts that a regulatory penalty for selling reserves just before a shortage will reduce forward volume when the market is expected to be tight.

4. Evidence of forward price premia

In this section, we examine the first of the theoretical implications, that forward prices should exceed expected spot prices when spot markets are expected to be tight. We do this by comparing forward prices to spot prices in local markets for natural gas. While a forward premium that is higher when spot markets are expected to be tight does not distinguish amongst the three models, it does establish that market behavior is inconsistent with the presumption of efficient forward markets that are unbiased predictors of spot prices. In section 5, we attempt to distinguish amongst the theories we have presented by examining data regarding forward trading quantities.

4.1 Local natural gas markets—data

Price data for local natural gas markets are sourced from Platts' GASdat product, and consist of location-specific observations in the day-ahead (spot) markets and the forward month ("bidweek") markets. Platts obtains prices in spot markets via surveys of trades made at each location and reports average location-specific spot prices at daily intervals.

Bidweek data occur on a monthly basis and represent the volume-weighted average price of all surveyed trades that occur at each location during bidweek. Bidweek takes place over the last five trading days of each month and consists of trades for gas to be delivered during the following month.[19]

Because our aim is to compare forward prices to spot prices, we must make the daily spot data compatible with the monthly bidweek data. We therefore average the daily spot observations within each location and month to obtain an average spot price for the month, and proceed at this level of aggregation for the remainder of the paper.[20] There exist 11,169 location-months for which both spot and bidweek prices exist,[21] spread over 131 locations across which the duration of coverage varies. For example, while data for Henry Hub in Louisiana span 1993 to 2007, data at the Carthage Hub in northeast Texas are only available for 1997 to 2002. These variations in data availability occur because trading activity in some locations varies over time, and Platts does not record observations when there are an insufficient number of trades to allow it to determine the average price.

Summary statistics for the spot and bidweek prices are shown in table 1. Both data series are highly right-skewed, as indicated by the excess of the mean over the median prices and by the large observed maximum prices. The summary statistics of the spot and bidweek prices are very similar, and the difference in means of 0.8 cents is not statistically significant.[22] Thus, on average, there is no statistically discernible forward premium or discount in prices for natural gas. This result, however, does not speak to whether forward premia develop when the market is expected to be tight, a topic to which we now turn.

4.2 Local natural gas markets—empirical framework

Our working hypothesis is that the forward price premium should rise when LDCs believe that they may be unable to obtain gas in the spot market due to supply constraints

[19] Platts will sometimes use the median of reported prices if it finds that one high-volume transaction skews its bidweek sample. Unfortunately, there are no indicators in the data to determine which observations are computed in this way.

[20] We use a straight average of the daily spot prices, rather than a volume-weighted average, because the straight average better reflects the nature of bidweek transactions, which specify a fixed volume of gas to be delivered every day of the month.

[21] The count of 11,169 location-months includes only those observations for which rolling regression spot price predictions can ultimately be generated, as discussed later. Without this restriction, there are 17,438 location-months.

[22] Statistical significance was tested via a paired t-test with standard errors clustered on year-month: the t-statistic is 0.15.

or coordination problems, a situation we have been calling a "tight" spot market. A natural measure of whether the spot market is expected to be tight is the expected spot price, because natural gas is most valuable precisely when access to it is difficult. Natural gas supply in local markets exhibits significant short-run constraints due to limited local storage and limited capacity to bring in additional gas by pipeline. In addition, the overall U.S. natural gas market can exhibit short-term supply constraints due to constraints on production rates at natural gas wells.[23] These factors lead the natural gas supply function in any period (distinct from the forward supply curve discussed in the previous section) to be relatively elastic up to the binding supply limits and then very inelastic at those limits, a familiar inverted L-shape supply curve. Thus, an expectation of a high spot market price (controlling for historical levels and recent trends) will coincide with an expectation that equilibrium is likely to occur on the inelastic portion of the supply curve, which is what happens when the market encounters a supply constraint—a tight market.[24]

To test our hypothesis that forward price premia will occur when spot markets are expected to be tight, we wish to estimate the parameters of the following equation:

$$BidWeek_{it} - Spot_{it} = \beta_0 + \beta_1 \; E[Spot_{it}] + \mu_i + g(t) + \epsilon_{it}. \quad [1]$$

Here, $BidWeek_{it}$ is the price during bidweek of month $t-1$ for gas to be delivered at location i in month t, and $Spot_{it}$ is the spot price for gas at location i in month t. $E[Spot_{it}]$ is the expectation at the beginning of bidweek in month $t-1$ of the spot price at location i in month t. The μ_i are location fixed effects and $g(t)$ is a 4th-order polynomial in time that controls for the secular upward trend in natural gas prices over the sample period. If forward price premia arise when the market is expected to be tight, then β_1 will be positive.

In order to estimate equation [1] directly, we would have to observe $E[Spot_{it}]$. Because we do not observe market participants' expectations of spot prices for a particular month, we must take an indirect approach. We begin by noting that markets for natural gas are markets in which the same participants operate month after month, that the participants are professional traders and purchasing agents, and that the transactions are for sizeable

[23] These constraints are indicated by increases in natural gas futures prices during the winter months; gas production is unable to fully respond to seasonal demand swings.

[24] At constrained times, a shift in demand due to weather or other exogenous shocks will also create larger changes in price than at unconstrained times. Thus, tight markets and high expected spot prices will also be associated with greater expected price volatility; *i.e.,* price risk.

monetary values. Under these circumstances, we expect that market participants' expectation of the spot price will be an unbiased predictor of the realized spot price. That is, market participants will incorporate all available information into their expectations so that any difference between their expectation and the realized spot price will be due to information revealed after expectations are formed. This information will therefore be orthogonal to the information that is incorporated into the expected spot price:

$$Spot_{it} = E[Spot_{it}] + \xi_{it}. \qquad [2]$$

Using equation [2], we can estimate equation [1] by substituting $Spot_{it}$ for $E[Spot_{it}]$ and accounting for the endogeneity introduced by ξ_{it}:

$$BidWeek_{it} - Spot_{it} = \beta_0 + \beta_1\ Spot_{it} + \mu_i + g(t) + (\epsilon_{it} - \beta_1 \xi_{it}). \qquad [3]$$

Using $Spot_{it}$ as a right-hand-side regressor introduces both classical measurement error and simultaneity bias, since the dependent variable is $BidWeek_{it} - Spot_{it}$. In order to use $Spot_{it}$ as a regressor, we need an instrument that is correlated with the realized spot price but uncorrelated with the error $(\epsilon_{it} - \beta_1 \xi_{it})$.

As an instrument, we construct a forecast of the spot market price that is based on information available to us and determined prior to the bidweek market in period $t-1$. To illustrate our approach, suppose we wish to forecast the spot price at the Chicago Citygate in November 2001, using information available to traders at the start of bidweek in October (October bidweek is when forward contracts are set for delivery in November). We cannot, of course, use the bidweek price itself to create the forecast, as our aim is to test for bias in the bidweek price during tight market periods. Instead, we construct the forecast using two additional pieces of information: (1) the NYMEX futures price of gas at Henry Hub, Louisiana, for delivery in November 2001, priced at the start of bidweek;[25] and (2) the spot price differential from Henry Hub to Chicago, also priced at the start of bidweek. The former yields a measure of the expected price of gas at Henry Hub in November, which will be correlated with November prices nationwide, while the latter measures the price

[25] A NYMEX (New York Mercantile Exchange) futures contract specifies a price for delivery of gas for a calendar month at the Henry Hub pipeline interconnect and storage facility in Louisiana. Because of the deep liquidity of both the NYMEX futures market and the Henry Hub spot market, and the fact that the vast majority of NYMEX market participants have no intention of taking physical delivery of gas, NYMEX futures prices are not subject to security of supply concerns that might drive a forward price premium in tight markets. Thus, we may use the NYMEX futures market to generate unbiased forecasts of Henry Hub spot prices.

differential to Chicago near the end of October, which will be correlated with the price differential in November.[26]

More generally, for all locations and months in our data, we forecast the expected spot price for location i in month t using the equation [4] below, in which $Future_{HH,t}$ is the NYMEX futures price for delivery at Henry Hub in month t, taken at the start of bidweek in month $t-1$; $Spot_{i,t-1}$ is the spot price at location i at the start of bidweek in month $t-1$; and the d_i are location fixed effects.

$$SpotForecast_{it} = a_0 + a_1\, Future_{HH,t} + a_2\, [Spot_{i,t-1} - Spot_{HenryHub,t-1}] + d_i \qquad [4]$$

To use equation [4] to predict expected spot prices, we must first estimate the equation's parameters—a_0, a_1, a_2, and the d_i—by running the regression specified in equation [5] below. This equation includes an unobserved orthogonal disturbance ν_{it} to account for information revealed between the start of bidweek in month $t-1$ and month t.

$$Spot_{it} = \alpha_0 + \alpha_1\, Future_{HH,t} + \alpha_2\, [Spot_{i,t-1} - Spot_{HenryHub,t-1}] + \delta_i + \nu_{it} \qquad [5]$$

In the process of estimating equation [5] and then generating forecasts using equation [4], we take care to avoid using any future information in our forecasts. That is, when we forecast the spot price for month t using equation [4], we only use parameters estimated using information revealed prior to month $t-1$. This means that we do not use our entire sample of spot price information to produce estimates of the α's and δ_i's from equation [5], and then apply these estimated parameters to generate a full time series of prices from equation [4].

Instead, we estimate equation [5] using a "rolling regression" approach. Rather than estimate a single set of $\hat{\alpha}$'s, we estimate a different coefficient vector $\hat{\alpha}_t = (\hat{\alpha}_{0t}, \hat{\alpha}_{1t}, \hat{\alpha}_{2t}, \hat{\delta}_t)$ for each month t using data only up to month $t-2$. These coefficients are then substituted for the corresponding a's and d in equation [4] to generate expected spot prices for month t. While this approach ensures that our spot price prediction for any month t does not include any information revealed after t, it does come with the cost that there are few data with which to estimate equation [5] in the early part of our sample. To avoid generating

[26] As an alternative to our instrumental variables approach, we could replace $E[Spot_{it}]$ in equation [1] directly with our forecast of spot price. However, doing so would cause the standard errors we estimate to be biased downwards (Murphy and Topel, 1985).

estimates based upon only a handful of points, we predict spot prices only for locations and months for which the coefficient vector in [5] was estimated using at least 24 observations.[27]

4.3 Local natural gas markets—results

We first report results from the estimation of equation [5]—the first stage regression we use to create forecasts of spot prices. For illustration, the first column of table 2 reports results using the full sample of spot price data. Our two primary determinants of spot price are the NYMEX future price and the spot price differential from Henry Hub to each location. The estimated coefficients of these variables are positive, as expected, and statistically significant with location fixed effects and standard errors clustered on year-month. The NYMEX futures price and the spot price differential clearly carry useful information with which to predict current prices.

Table 2 also reports the results of rolling regressions of equation [5]. We ran 174 rolling regressions to generate parameters for the forecast of spot prices from February 1993 to July 2007. The estimated coefficient on the NYMEX future price is positive in every regression, and the estimated coefficient on the price differential is positive for 168 of 174. We use these estimated rolling regression parameters to forecast spot prices using equation [4], and then use these forecasts as instruments for spot price in our main specification, equation [3].

The results of estimating equation [3] are reported in table 3. The results presented in column 1 are those obtained through the estimation of [3] as written. Column 2 adds month-of-year fixed effects, and column 3 adds interactions between location fixed effects and a linear time trend. Across all three specifications, we estimate that the forward premium of bidweek prices over spot prices increases systematically with the expected spot price. The point estimates indicate that a \$1.00 rise in the expected spot price (as measured by the instrumented spot price) is predicted to cause a \$0.20 to \$0.21 rise in the forward premium, depending on the empirical specification.

The p-values for the results in columns 1 through 3 are 0.094, 0.066, and 0.064, respectively, with standard errors that are clustered on year-month.[28] While clustering the

[27] There is a tradeoff in setting the minimum number of observations required to estimate equation [5]: a high minimum number yields more precise predictions of spot prices, but reduces the total number of predictions and therefore reduces the number of observations in the main specification, equation [3]. The primary results concerning forward price premia and trading quantities in local natural gas markets do not qualitatively change as we adjust the minimum number, even when we increase it to 48 months.

[28] Clustering instead on location, or not clustering at all, yields much smaller standard errors, under which the results are significant at the 1% level.

standard errors on year-month addresses correlation of the disturbances across locations, it does not address inter-temporal correlation within a location. We therefore re-run specifications 1 through 3 using Prais-Winsten GLS, and report these results in columns 4 through 6. The estimated autocorrelation parameters are small, and the GLS coefficients and standard errors differ little from the IV estimates in columns 1 through 3.

These results provide evidence of forward price premia in local natural gas markets at times and locations for which these markets are expected to be tight. This finding suggests that natural gas markets are not frictionless, liquid, riskless markets. At a minimum, when markets are expected to be tight, arbitrage between forward and spot markets fails to bring forward prices down to expected spot prices. In section 3, we modeled several factors that could lead to this forward premium, including security of supply, excess caution, and price risk aversion. While a forward price premium is consistent with all three of these explanations, only excess caution is likely to lead to a reduced volume of transactions when markets are expected to be tight. To resolve these models, we next analyze data on forward transaction volumes.

5. Evidence of reduced forward quantities

While "excess caution," brought about by an asymmetric regulatory penalty for stocking out of gas following a forward sale rather than a forward purchase, can explain the observed forward price premia, other theories of behavior in these markets are also consistent with this evidence. As discussed in section 3, appropriately shaped demand curves for gas, reflective of security of supply concerns, may also cause forward premia when markets are expected to be tight, as can price risk aversion on the part of of LDCs. However, only a model of the excess caution of LDCs in selling gas predicts that forward trading quantities will decrease in spot markets that are expected to be tight. This reduction occurs because the number of potential sellers decreases as fewer firms are *ex ante* certain that they have adequate gas to meet their maximum possible demand. On the other hand, both a model of security of supply without excess caution and a model of price risk aversion suggest that trading volumes are likely to rise when spot markets are expected to be tight; neither of these alternative models can convincingly explain a significant decline in forward trading volume in such markets. We may therefore distinguish amongst these theories, or at least assess which is dominant, by examining data on forward quantities in markets that are expected to be tight or expected to be loose.

In addition to price data, Platts also reports the volume of gas traded during bidweek at each market location from June 1999 to July 2007, with a gap in coverage from July

2002 to June 2004.[29] We examine whether bidweek volumes are relatively low at locations and months in which the market is expected to be tight, following an empirical strategy paralleling that used to examine forward price premia. Specifically, we wish to estimate the following equation, which uses expected spot price as the measure of market tightness and controls for location fixed effects and a flexible time trend.

$$\ln(Volume_{it}) = \gamma_0 + \gamma_1 \ E[Spot_{it}] + \mu_i + g(t) + \eta_{it}. \qquad [6]$$

We encounter the same problem in estimating equation [6] as we faced in estimating equation [1]: we do not observe the expected spot price. We solve this problem using the same solution that we used to estimate forward price premia in tight markets. Because the realized spot price is equal to the expected spot price plus white noise, per equation [2], we may re-write equation [6] as equation [7]below. We estimate equation [7] by instrumenting for the realized spot price using the same *ex ante* price forecast—based on equations [4] and [5]—that was used as an instrument in equation [3].

$$\ln(Volume_{it}) = \gamma_0 + \gamma_1 \ Spot_{it} + \mu_i + g(t) + (\eta_{it} - \gamma_1 \xi_{it}). \qquad [7]$$

Equation [7] takes as the dependent variable the natural logarithm of the total volume of gas (measured in millions of cubic feet per month) traded in market i during bidweek of month $t-1$ for delivery in month t (note that in designating this variable as $Volume_{it}$, the t refers to the delivery month). We use the logarithm of volume to avoid scaling problems associated with the fact that some locations generally see much larger volumes than others: average volumes at each location range from 4 million cubic feet per month to nearly 1,400 million cubic feet per month.

Matching the bidweek volume data to the spot price forecasts yields a sample with 5,062 observations, spread over 108 locations. Full summary statistics for bidweek volume are reported in the final row of table 1.

Column 1 of table 4 reports the results of estimating equation [7] with the bidweek volume data. The estimated coefficients demonstrate that forward volumes decline significantly in tight markets. A $1.00 increase in the expected spot price is predicted to decrease the logarithm of forward trading volume by 0.079, equivalent to a decrease in volume of

[29] This gap occurs due to a changeover in publications by Platts, prompted by its merger with FT Energy in 2001. While prices are reported for these dates, volumes are not. In addition, there are a further 183 observations outside of these dates during which prices are reported, but not volumes.

7.6%. The effect is statistically significant at the 1% level and is robust to adding month fixed effects (column 2) and interactions between location fixed effects and a linear time trend (column 3). These latter results indicate that the reduction in forward volume in response to high expected prices is not merely a seasonal phenomenon, nor is it driven by long-term secular trends.

As with estimating the forward price premium equation [3], one might be concerned that the residuals from estimating equation [7] are autocorrelated. In columns 4 through 6, we estimate equation [7] using Prais-Winsten GLS. There is much more evidence of autocorrelation here than in the price regression; the estimated autocorrelation coefficients ($\hat{\rho}$) range from 0.34 to 0.54, depending on the specification. The estimates of the coefficient of interest—the effect of the expected spot price on bidweek volumes—are slightly higher when estimated using GLS than with regular IV. The result presented in column 4, for example, indicates that a $1.00 increase in the expected spot price leads to a decrease in forward volume of 9.4% on average—a larger magnitude than that reported in column 1. The estimated decreases in volume are still statistically significant under GLS: the p-values of the results reported in columns 4 through 6 are 0.032, 0.012, and 0.00008, respectively.

The implication of these results are that trade in forward markets is reduced when spot prices are expected to be high. If, as we have argued, high spot prices coincide with tight spot markets, then this suggests that there is a market inefficiency at work: forward markets shrink at the very time they are most needed, namely when they could compensate for tight spot markets. These reductions in forward quantities traded in tight markets are consistent with a model in which the regulator penalizes curtailments caused by forward sales more than curtailments caused by insufficient forward purchases. Absent the "excessively cautious" behavior caused by this regulatory asymmetry, a model of security of supply concerns would predict the opposite of this result, as would a model of price-risk aversion.[30]

6. Supplementary evidence from markets for gas transportation

The evidence we have presented so far has concerned the effect of tight — markets

[30] We also investigated regional variation in the price and quantity effects that we have estimated. The Northeast area (New York, New Jersey, and New England) is the area that one might expect to see the strongest effects, due to the combination of large weather variation and limited pipeline delivery capacity. These conditions are reflected by the fact that the Northeast has both the largest average price differential to Henry Hub of any region as well as the largest variance in this differential. We found that the estimated coefficients on the forward price premium (table 3) were twice as large for the Northeast as for all locations, and statistically distinct from the average effect at the 5% level. The estimated impact on bidweek volume, however, (table 4) was about the same for the Northeast as for all locations.

for natural gas on the prices and volumes of transactions in local forward markets. As we noted in section 2, an LDC can obtain the gas it needs to serve local market demand either by buying gas in its local market, or by buying gas in an upstream, producing market and then arranging for transportation from the producing region to the local market. We now investigate the effect that tight spot markets for natural gas have on markets for pipeline transportation. We expect that the anticipation of tight spot markets would lead to lower volumes in transportation markets, just as we found they did in natural gas markets. Because data coverage is not as extensive for the transportation market as it is for the natural gas markets we examined above, we categorize this section as supplementary evidence.

6.1 Markets for pipeline capacity

Pipelines sell rights for most of their capacity to shippers under long-term contracts that can be as long as 30 years in length. There is a maximum price set by FERC, known as the "reservation charge," that shippers can be charged for this property right.[31] A shipper who holds long-term transportation rights can sell part or all of its rights to another shipper for a fixed period of time, usually one calendar month, in a secondary market called the capacity release market. The price for this transfer is negotiated between the "releasing" and "awarded" shippers. Except for a brief "waiver period" from April 2001 through September 2002, the capacity release transfer price was capped at the maximum reservation charge set by FERC.[32] Because the price cap on the capacity release market effectively constrains trade in tight markets—presenting a significant confound in our analysis—we only examine releases that occur during those months in which the cap was waived.

The Platts dataset that provided us with natural gas prices and volumes at locations throughout the U.S. also contains information on transactions in the capacity release market. In these data, each observation represents a transaction in which capacity rights are sold for a fixed time period on a specific pipeline route. Unlike the spot and bidweek gas market data, the capacity release data are not based on surveys; they are a comprehensive set of all releases that have occurred on each pipeline since Platts began tracking them.[33] Each observation provides information regarding the transaction date, the duration of the

[31] The charge is set at a level that will recover pipeline capital costs plus a reasonable rate of return on investment.

[32] The FERC issued the temporary waiver for the price cap, in Order 637, because it was concerned that the cap constrained trade in tight markets. However, it let the waiver expire without comment at the end of September 2002.

[33] The year in which observations begin varies by pipeline, but is generally in the late 1990s.

release, and the route along which capacity is released (many pipelines have multiple routes either because they have a branched structure or because there exist receipt and delivery points at intermediate locations along the line).

In the capacity release market, the expected spot value of the capacity is the expected spot price differential between the two endpoints of the capacity release route. We call this differential the *spot margin*. We expect high spot margins to coincide with high spot prices for gas in downstream markets, because downstream spot markets will be tight only when the capacity to transport gas into those downstream markets is fully utilized. It is during these capacity constrained periods that the differential between upstream and downstream gas prices may be large.

To facilitate the linking of capacity release transactions to their corresponding expected spot margins, we define a capacity release *product* as a pipeline-route-month combination for which capacity may be released. For example, capacity on Transcontinental Pipeline's route from Louisiana to New York for October 2003 constitutes a product. For each product for which the necessary data are available, we compute a *spot margin* by subtracting the spot price at the upstream node of the route during the relevant month from the spot price at the downstream node. We also subtract the small variable cost of pipeline transportation that must be paid to the pipeline by any user of capacity rights. This margin represents the profit that a shipping firm would make, absent the capacity release payment, by shipping gas on the released route over the contract term, and is therefore an *ex post* measure of the value of a capacity release for that product. Each product is associated with exactly one spot margin; however, for any given product we may observe zero, one, or multiple capacity releases.[34]

There are two types of capacity releases: recallable and non-recallable. In a recallable release, the releasing firm has the right to recall its capacity from the awarded firm after a contractually agreed notice period that can be as brief as 24 hours. The distinction between recallable and non-recallable releases is important to our analysis. We aim to investigate the reluctance of LDCs to sell forward capacity in tight markets at times when they may later need the capacity to ship gas to serve demand. Recallable releases should be less affected by excess caution than non-recallable releases: recall rights allow an LDC to recover its capacity in the event that it is needed (though a lag may be involved).

[34] Some capacity release observations are dropped from the dataset for institutional reasons. We drop releases that occur between companies that are affiliates, a small number of releases that are for terms other than one calendar month, and releases for which the transaction occurs more than one month ahead of the effective month of the release.

We therefore divide our analysis of capacity releases into recallable and non-recallable releases. We anticipate that non-recallable releases will be significantly less frequent in tight markets, while the frequency of recallable releases will be relatively unaffected by market tightness.

Our data include capacity releases on 11 routes on five major interstate pipelines: Columbia Gulf, Florida Gas Transmission, Texas Eastern, Texas Gas Transmission, and Transcontinental.[35] The geographic distribution of these pipelines is indicated in the map in figure 12. While all five pipelines receive gas in the Texas and Lousiana production basin, their delivery points vary, and include New York, Appalachia, and Florida. As shown in table 5, the final capacity release dataset contains 155 products, 238 non-recallable capacity releases, and 456 recallable releases. These are not spread uniformly across the five pipelines or over the 18 months of observation. Eighty-seven products have no observed non-recallable releases, whereas there are 3 products for which we observe more than ten. Similarly, there exist 50 products that have no recallable releases, yet 4 products have more than ten.

The distribution of spot margins is significantly right-skewed, consistent with a gas pipeline infrastructure that has spare capacity at most times, but occasionally does become constrained.[36] It is during these constrained periods that possession of firm transportation rights can be extremely valuable.

6.2 Capacity release markets estimation and results

In this section, we examine the frequency with which we observe capacity releases and assess whether transactions are less frequent when the market is expected to be tight (that is, when spot margins are expected to be high). We measure the frequency of capacity release transactions as the number of transactions during month $t-1$ for capacity on pipeline route i during month t.

$$CR\ Transactions_{it} = \theta_0 + \theta_1\ E[SpotMargin_{it}] + \mu_i + g(t) + \zeta_{it}. \qquad [8]$$

In estimating this specification, we have the same problem that was encountered in estimating equations [1] and [6]: we do not observe the expected spot margin. Similar to

---

[35] While capacity release data were available for several other major pipelines, no more than thirty transactions were observed for those lines, and they were therefore dropped.

[36] The median spot margin is \$0.07, the mean is \$0.13, and the maximum is \$1.61.

the approach used previously, we note that the realized spot margin should be equal to the expected spot margin plus an orthogonal error:

$$SpotMargin_{it} = E[SpotMargin_{it}] + \upsilon_{it}. \quad [9]$$

When we substitute equation [9] into equation [8] we obtain:

$$CR\ Transactions_{it} = \theta_0 + \theta_1\ SpotMargin_{it} + \mu_i + g(t) + (\zeta_{it} - \theta_1 \upsilon_{it}). \quad [10]$$

In previous specifications, we used an instrumental variable to accommodate the measurement error in our explanatory variable. However, the instruments that we used previously are not suitable here: the NYMEX futures price for Henry Hub will be correlated with the levels of prices throughout the country, but will not be correlated with the differential between prices at different locations. In light of this, we proceed by estimating equation [10] directly, noting that attenuation bias will work against the results we find.

Given the integer nature of the number of releases per product, as well as the presence of many products without any releases, we estimate equation [10] using a Poisson count model. We report estimation results as incidence rate ratios (IRR's). Each IRR should be interpreted as the proportional change in the rate of occurrence of capacity releases as the result of a $1.00 increase in the expected spot margin. The "null" effect is an IRR of one, and the p-values we report are for a test with this null hypothesis (standard errors are clustered on year-month). An IRR of more than one indicates that an increase in expected spot margin tends to increase the number of capacity releases, and an IRR of less than one indicates the opposite. The attenuation bias noted above will bias the estimated IRR towards one.

Results for non-recallable releases are reported in table 6. The interpretation of the estimated IRR of 0.1149 indicated in column 1 is that a $0.10 increase in expected spot margin will decrease the number of capacity release transactions by 19.5% on average. The p-value of this effect against the null is 0.053. When a time trend is added to the specification, as shown in column 2, the estimated coefficient is similar, and the statistical significance slightly decreases. When month fixed effects are added to the specification, as shown in column 3, the estimated effect decreases in magnitude (an increase in the expected spot margin of $0.10 would decrease the number of transactions by 13.1%) and the p-value increases to 0.155. Overall, these results buttress the evidence obtained from local natural gas markets that forward trades are reduced when spot markets are expected to be tight.

Table 7 reports the results of estimating the Poisson model with recallable capacity releases. As expected, the estimated effect of tight markets on the number of recallable releases is substantially weaker than that reported for non-recallable releases. Across all three specifications, the estimated IRR's are much closer to one and not statistically significant. Due to the imprecision of the estimates, however, pooling of the estimates for recallable and non-recallable capacity releases cannot be rejected.

The analysis of capacity release transactions supports the existence of an asymmetric regulatory penalty for selling reserves in the forward market and then curtailing customers. There is an economically and statistically significant decline in the number of non-recallable capacity releases when the market is expected to be tight. Furthermore, there is much less evidence of such an effect for recallable capacity releases, for which the regulatory penalty is less salient.

7. Conclusion and Broader Relevance

The natural gas industry involves firms that face security of supply concerns regarding the procurement of essential inputs. When spot markets for these inputs are thin—as is often the case in natural gas—then firms may be willing to pay a price premium to purchase inputs on forward rather than spot markets, so as to guard against supply shortages. Furthermore, the regulatory environment punishes particularly harshly those delivery curtailments that can be linked to a utility having made forward sales of gas. We show theoretically and empirically that the excess caution these regulatory incentives impose on utilities can distort forward natural gas markets when demand is expected to be high. At such times, this caution will cause forward prices to exceed expected spot prices, and cause volume in the forward market to decline. In particular, there will be a reduction in efficient trade at precisely those times at which trade is most needed: when demand is expected to be high and social surplus gain is greatest from transferring gas to those who value it most.

As principal-agent incentive problems and security of supply concerns are hardly unique to the natural gas industry, we suspect that these considerations may exist in other industries as well. For instance, consider the hypothetical example of a manager of an auto plant, who has on hand more tires than he thinks he is likely to need given the range of output levels the plant might be called on to produce in the next month. Should he try to rid the plant of some of the excess supply of tires if he can do so at a profit? If he makes a profit on the sale of the tires, it is likely to lead to a small reward. If, however, the plant is called upon to produce an unusually large number of cars, and the production line has to be shut down because there aren't enough tires, then the manager is likely to suffer substantial negative consequences. In this case, organizational incentives rather than regulatory incentives are the root cause of the manager's caution. However, in both this case and the case of natural gas, a forward sale of necessary inputs by the agent can be used by the principal to infer that agent is lacking in either effort or ability, and then to punish the agent severely, perhaps even by firing him and replacing him with a new agent.

The principal-agent model presented in appendix 2 indicates that these penalties for observable actions that can be linked to negative outcomes may be efficient from the principal's point of view, suggesting that rational managers in an organization may be willing to apply them. However, this paper demonstrates that the impacts of these incentives extend beyond the principal-agent relationship in which they are applied, and can cause significant market inefficiencies. In markets for necessary inputs, the excess caution created by these principal-agent relationships can cause firms to be reluctant to engage in trades, even when differences in marginal valuations are significant.

References

Brandenburger, Adam and Ben Polak. “When Managers Cover their Posteriors: Making Decisions the Market Wants to See,” *RAND Journal of Economics*, **27**(3)(Autumn 1996), pp. 523-541.

Bodie, Zvi and Victor I. Rosansky. “Risk and Return in Commodity Futures,” *Financial Analysis Journal*, **36**(May/June 1980), pp. 27-39.

Carlton, Dennis W. “Contracts, Price Rigidity, and Market Equilibrium,” *The Journal of Political Economy*, **87**(5) (Oct., 1979), pp. 1034-1062.

Carlton, Dennis W. “The Rigidity of Prices,” *The American Economic Review*, **76**(4) (Sep., 1986), pp. 637-658.

Carter, Colin A., Gordon C. Rausser and Andrew Schmitz. “Efficient Asset Portfolios and the Theory of Normal Backwardation,” *The Journal of Political Economy*, **91**(2) (April 1983), pp. 319-331.

Chambers, Marcus J. and Roy E. Bailey. “A Theory of Commodity Price Fluctuations,” *The Journal of Political Economy*, **104**(5) (Oct., 1996), pp. 924-957.

Cremer, Helmuth, Farid Gasmi, and Jean-Jacques Laffont. “Access to Pipelines in Competitive Gas Markets,” *Journal of Regulatory Economics*, **24**(1) (2003), pp. 5-33.

Deaton, Angus and Guy Laroque. “Competitive Storage and Commodity Price Dynamics,” *The Journal of Political Economy*, **104**(5) (Oct., 1996), pp. 896-923.

Deaton, Angus and Guy Laroque. “On the Behaviour of Commodity Prices,” *The Review of Economic Studies*, **59**(1) (Jan., 1992), pp. 1-23.

Dusak, Katherine. “Futures Trading and Investor Returns: An Investigation of Commodity Market Risk Premiums,” *The Journal of Political Economy*, **81**(6) (Nov.-Dec., 1973), pp. 1387-1406.

Finoff, David, Curtis Cramer, and Sherrill Shaffer. “The Financial and Operational Impacts of FERC Order 636 on the Interstate Natural Gas Pipeline Industry,” *Journal of Regulatory Economics*, **25**(3) (2004), pp. 243-270.

Harbaugh, Richmond. “Prospect Theory or Skill Signalling?” Working paper, Indiana University (2006).

Holmstrom, Bengt. “Managerial Incentive Problems: A Dynamic Perspective,” *Review of Economic Studies*, **66**(1) (1999), pp. 169-182.

Lyon, Thomas P. “Regulation with ‘20-20 Hindsight’: Least Cost Rules and Variable Costs,” *Journal of Industrial Economics*, **40**(3) (Sep., 1992), pp. 277-289.

Murphy, Kevin M. and Robert H. Topel. “Estimation and Inference in Two-Step Econometric Models,” *Journal of Business and Economic Statistics*, **3** (1985), pp. 370-379.

Palley, Thomas I. “Safety in numbers: A Model of Managerial Herd Behavior,” *Journal of Economic Behavior and Organization*, **28**(3) (1995), pp. 443-450.

Scharfstein, David S. and Jeremy C. Stein. "Herd Behavior and Investment," *The American Economic Review*, **80**(3) (Jun., 1990), pp. 465-479.

Williams, Jeffrey C. and Brian D. Wright. *Storage and Commodity Markets.* New York: Cambridge University Press, 1991.

Appendix 1: Forward Market Volume and Price Premium Under Security of Supply Concerns

Consider a model in which several wholesale gas buyers—LDCs who need to meet end-use customer demand—are present at the downstream end of a pipeline. The LDCs holds "reserves" in either local storage or through firm contracts for pipeline delivery at the time purchasing for a given period begins. Customer demand varies exogenously due to weather and other shocks and is extremely inelastic in the short run.

When the period of delivery arrives, the exogenously-determined customer demand and the storability of excess supply combine to yield a convex LDC marginal value function. The marginal value of supplies is very high for units up to the customer demand that the LDC faces and flattens out rapidly for units beyond that level, because additional quantities can be put into storage. At the delivery period the LDC faces very asymmetric costs of failing to match its procured supply with customer demand—much greater costs if it has too little supply than if it has too much. If, however, it could always be sure that it could adjust its total procurement quantity to exactly match its customers' demand through use of the spot market, this asymmetry would be of no consequence.

Consider the situation that the LDC faces when it attempts to buy its requirements before retail demand shocks are known, *i.e.,* the extreme case in which no adjustments are feasible after demand shocks are revealed, because there are no spot market transactions.

Assume that an LDC knows its inverse marginal value function, $D(q, w)$, and assume that the probability distribution of the weather and other demand shocks that could occur, $f(w)$, is common knowledge. Prior to w being revealed, the LDC knows that it could have one of many different demand functions based on different draws of w. For the reasons just discussed, assume $D_q(q, w) < 0$, $D_{qq}(q, w) > 0$, and $D_w(q, w) > 0$, where the subscripts denote derivatives.

Assume that the LDC is risk neutral with respect to price. Thus, recognizing that it can make no adjustments in a spot market, the LDC would purchase or sell gas in the forward market according to the demand function $\bar{D}(q)$ that would be the weighted average demand function based on the probability distribution of $f(w)$, $\bar{D}(q) = \int_w D(q, w) f(w) dw$. That is, $\bar{D}(q)$ is the probability-weighted average of the marginal valuation of additional gas at a given q over all possible demand states. Since $\bar{D}(q)$ is a weighted average of convex functions, with non-negative weights, it follows that $\bar{D}(q)$ is convex.

Result 1: Define $\hat{q}$ as the LDC's expected spot market demand at spot price p, and assume its inverse demand curve satisfies the sufficient condition: $\forall(q, w)$ such that $D(q_i, w_i) = D(q_j, w_j)$ and $w_i < w_j$, we have $D_q(q_i, w_i) \geq D_q(q_j, w_j)$. Then $\bar{D}(\hat{q}) > p$. That is, the firm's marginal forward valuation at the quantity $\hat{q}$ that is its expected demand at spot price p is greater than p. Equivalently, for any forward market price p, an LDC will choose to purchase a larger quantity in the forward market than it would in expectation if it faced that same p in the spot market (*i.e.,* after w is revealed).

The result follows immediately from the convexity of $D(q, w)$ in q and the condition that a positive weather shock causes demand to become (weakly) steeper at every given price. Essentially, the value of holding extra supplies of gas before w is revealed is large because the cost of a shortage (moving leftward up a steep demand curve) is much greater than the cost of having excess supplies (moving rightward down a flatter demand curve) at delivery time.

The sufficient condition in result 1 is plausible, but not general. Parallel shifts of the (convex) demand curve would result in such a forward demand premium. Proportional quantity increases at all prices, however, need not; at a given price, a participant's demand in the forward market could be greater or less than its expected spot demand at that price. If result 1 does hold, then it is straightforward to show that the equilibrium forward price could be greater than the expected spot price.[37]

In reality, of course, a spot market does exist and some transactions occur after w is revealed. There is some risk, however, that a firm will not be able to access the spot market for some trades due to lack of market liquidity. Assume that the probability a firm will be unable to access the spot market is ϕ.

If the firm could access the spot market with certainty ($\phi = 0$), then, because it is risk neutral with respect to price, it would be unwilling to buy in the forward market at any price above the expected spot price, which we will call $\bar{p}$, or sell at any price below the expected spot price. Its forward market demand curve would be flat at $\bar{p}$. If, however, a firm has some risk of failing to make a spot purchase, that failure has an expected cost in lost value equal to the value of the incremental reserves, $\bar{D}(q)$. Thus, a firm expecting a spot price of $\bar{p}$ will have a willingness to pay in the forward market of $\hat{D}(q) = (1-\phi)\cdot\bar{p}+\phi\cdot\bar{D}(q)$.

[37] A simple example in which a forward premium exists can be shown using figure 5. If there are two firms that each in any period draw either the w_1 or the w_2 demand functions, their draws are perfectly *negatively* correlated, and the total quantity available in the market is $2 \cdot E[q]$, then the expected spot price would be $\bar{p}$ and the futures price would be $\bar{D}(E[q])$, which would be greater than $\bar{p}$ under the conditions of result 1. An alternative example also illustrates one case is which we know there would be no forward premium or discount: if the demands of the two firms were perfectly correlated, then the expected spot price would just be the weighted average of $D(E[q], w_1)$ and $D(E[q], w_2)$, which would exactly equal each firm's $\bar{D}(E[q])$. We believe, but have not proven, that this is a special case and that if the demand shocks faced by the LDCs in the market are not perfectly rank correlated, then the forward price will be greater than the expected spot price.

The liquidity risk in the spot market means that the firm's marginal willingness to buy or sell in the forward market is a weighted average of its marginal forward valuation of reserves (with no access to a spot market) and the expected spot price. Access to a spot market with probability $1 - \phi$ reduces the magnitude of the forward valuation premium or discount, but does not eliminate it. This is illustrated in figure 6.[38]

While $\hat{D}(q)$ would be an LDC's demand for gas if it had no reserves going into the forward market, each LDC actually starts with an endowment of gas reserves, q_{res}, for which it has already contracted. It would supply quantity to the market if p is above its marginal forward marginal valuation at its current reserve level, denoted by $\bar{D}(q_{res})$, and it would purchase from the market if $p < \bar{D}(q_{res})$. With firms possessing different $\hat{D}(q)$ and different q_{res}, each would have a different $\bar{D}(q_{res})$ and different supply and demand curves, as illustrated in figure 4.

The supply and demand functions of all market participants aggregate into the market supply and demand, which then determine the equilibrium forward price and quantity. Besides the relationship between the forward and expected spot price, we are also interested in the forward quantity and how that changes with market conditions. To examine this relationship, we note first that volume in the forward market is the sum of the difference for each market participant between the reserves it holds going into the forward market and the quantity it would choose to hold at the equilibrium price. That is, forward market volume is equal to the mismatch between reserves and demand, $\frac{1}{2}\sum_m |q_{res} - \hat{D}^{-1}(p_e)|$, where m indexes market participants and the $\frac{1}{2}$ recognizes that each trade has one buyer and one seller. Volume would be expected to vary with the anticipated "tightness" of the gas market — which we measure as the expected spot price of gas — only to the extent that tightness is correlated with size of mismatches in allocation of reserves.

It might seem at first that no particular correlation would be expected, positive or negative, between market expected tightness and the size of reserve misallocations. However, we sketch three reasons that there seems likely to be a positive correlation, suggesting that this theory would imply forward volume would increase at times when the spot market is

[38] The forward premium in a tight-market scenario at first appears to create an opportunity for speculators to engage in a risky arbitrage that is profitable in expectation. However, the same liquidity risk that creates the premium also undermines the opportunity for profitable arbitrage. In order to take advantage of the forward premium, a speculator would have to sell gas in the forward market, and then buy those supplies from the spot market for delivery to the buyer from the forward market. Like other market participants, however, the speculator would face the liquidity risk that it could not obtain the supplies from the spot market that it had committed to deliver. In the event of a failure to deliver, the speculator would be liable for the damage imposed on the buyer of the forward gas, which would presumably reflect the buyer's realized $D(q, w)$. Thus, the speculator would be in no better position to take advantage of this premium than other market participants.

expected to be tight.

First is a simple scale effect. High prices and tight gas markets are associated with positive demand shocks much more than with negative supply shocks. Gas use is greater in the winter when prices are on average higher. If reserve misallocation increases with the total size of the reserve "pool," this would tend to manifest as a positive correlation between market tightness and forward volume.

Second is a heterogeneous impact effect. One would expect there to be more reserve misallocation at the time of the forward market if there is generally more uncertainty about demand levels for the relevant period. Uncertainty about demand that exists prior to the forward market, but is partially or completely resolved by the time of the forward market, will increase the level of misallocation that can be corrected through trading in the forward market. Not only are demand levels greater at times when prices tend to be higher, demand variability is also greater at those times. This is both *ex ante* because demand variability is greater in the winter and *ex post* because tight local markets generally reflect some weather shock, which is likely to have unforecastably heterogeneous effects on among market participants. Thus, misallocation is likely to be greater than in tight markets than in loose markets which are generally associated with less local demand uncertainty.

Third is an effect from overcoming transaction costs. For a given quantity of misallocated reserves, the gains from reallocation are likely to be greater if there is a tight market. In a loose market, the supply and demand curves in the forward market are fairly flat. Few or no participants are desperate to purchase reserves and the gains from most potential trades are small. Thus, if there is a per-unit transaction cost of trading reserves it will impede more trades during loose markets and volume will be lower. In contrast, if the market is expected to be tight, more participants are likely to have steep marginal valuation curves around their current quantity levels. In that case, the gains from trade will be greater and the transaction cost will impede fewer trades or, put differently, more of the potential trades will clear the transaction cost hurdle necessary for a trade to be worthwhile.

None of these arguments settles the question, but each tilts clearly in the direction of greater forward volume at times of tight gas markets. The magnitude of the effects is less clear. Still, it seems extremely unlikely that this model would explain *decreases* in forward market volume associated with tight gas markets.

Appendix 2: Principal-Agent Model of Procurement

Here, we offer a model and numerical example to demonstrate that an optimizing principal may benefit from the imposition of incentives that cause procurement agents to exercise "excess caution" in their sales of reserves. The primary features of the model correspond to those discussed in the paper: the principal's cost of being short of reserves is higher than the cost of holding excess reserves, and the agents are heterogeneous in ability, so that some are more skilled at forecasting demand than others. We show that the principal may want to fire an agent that held an insufficient level of reserves relative to *ex post* needs, but only if that agent had sold reserves prior to the shortage. The intuition driving the model is tied to the concept of "plausible deniability," in that the principal will not fire an agent who can plausibly claim that he held insufficient reserves only because there was no additional supply available for purchase in the market. While this asymmetric punishment does allow the principal to cleanly distinguish high-ability agents from low-ability agents, it leads both types of agents to be less likely to sell reserves.

Set-up

The timing of the model is as follows:

Period 0: The principal and the agent both observe the initial level of reserves, R_0.

Period 1: The agent receives a signal, S, about the probability distribution of the ultimate level of need, N, and decides on his target level of reserves, R^*.

Period 2: If $R_0 > R^*$, the agent sells reserves until he is left with R^*. If $R_0 < R^*$, the agent attempts to buy reserves, which may or may not be successful. The final level of reserves is R_2. After this, the principal observes $R_2 - R_0$, i.e., whether the agent bought or sold reserves.

Period 3: The level of need, N, is realized. The principal observes whether R_2 is greater than or less than N, and rewards or punishes the agent accordingly.

The level of need, N, is distributed uniformly on one of two intervals, either $[\mu_1 - r, \mu_1 + r]$ or $[\mu_2 - r, \mu_2 + r]$. The probability that N is distributed on either interval is equal to 1/2. All agents are aware of these ex-ante distributions.

Each agent may be either of the good type or the bad type. An agent is good with probability q and bad with probability $1 - q$. The good agent receives a perfect signal of which of the two possible distributions for N is correct; that is, he knows whether N is distributed on $[\mu_1 - r, \mu_1 + r]$ or $[\mu_2 - r, \mu_2 + r]$. The bad agent receives only a noisy signal of which distribution is the true one—specifically, the signal indicates which of the two distributions is correct with probability $s \geq 0.5$. If $s = 0.5$, the signal is useless, and if $s = 1.0$, a bad agent is equivalent to a good agent.

The agents know their types and are rational. In particular, the bad agent knows that his signal may be incorrect, and takes this into account in his decisions.

In period 2, if the agent wishes to sell, he can do so. If he wishes to buy, he may or may not be able to do so due to a thin marketplace. We model this by supposing that Y, a random variable, is the maximum amount that the agent will be able to buy if he decides to buy. Y is distributed according to the cdf $\Pr(Y \leq y) = 1 - e^{-\tau y}$, $y \geq 0$ so that the probability the agent will be able to buy at least y is given by $e^{-\tau y}$. The probability that the agent will be able to buy the quantity sufficient to reach his target level of reserves, R^*, is therefore $e^{-\tau(R^* - R_0)}$.

The principal's payoff function is given by:

$$
\begin{aligned}
U_P &= -F(R_2 - N) \text{ if } R_2 > N && (1)\\
&= -G(N - R_2) \text{ if } R_2 < N && (2)
\end{aligned}
$$

where $F(\cdot)$ and $G(\cdot)$ are both positive and monotonically increasing, $F(0) = G(0) = 0$, and $G(x) > F(x)\, \forall x$. This means that the principal's payoff function is maximized (at 0) when $R_2 = N$, and that shortfalls and excesses have asymmetric costs: a shortfall in reserves of any amount is more costly to the principal than excess reserves of the same amount.

While we model this problem as a single three period game, we recognize that a more comprehensive approach would consider a repeated game. In such a framework, the agents' valuations of future rounds would be modeled endogenously, and we would model the fact that, when the principal and agent interact repeatedly, the principal may gradually learn about the agent's type even if the agent does not sell reserves and come up short. Here, we abstract from the repeated game by compressing the model into one period, yet retain the features that cause an agent who sells reserves and then fails to serve demand to be fired.

The mechanism

Let the principal's transfers to the agent be defined as follows, in which W is a fixed wage (or more precisely, a wage in excess of what the agent could earn through outside employment), and α is a constant between zero and one:

$$\begin{aligned} A_P &= W - \alpha F(R_2 - N) \text{ if } R_2 > N && (3) \\ &= W - \alpha G(N - R_2) \text{ if } R_2 < N \text{ and } R_2 \geq R_0 && (4) \\ &= -\alpha G(N - R_2) \text{ if } R_2 < N \text{ and } R_2 < R_0 && (5) \end{aligned}$$

Under this mechanism, the agent is fired (and is no longer paid the fixed wage) if he is short of reserves after having sold reserves, but is not fired if he is short following a purchase of reserves or no change in reserve levels. In this sense, W can be thought of as the value of all future wages, to be received by the agent only if he is not fired.

The attractiveness of this feature to the principal will hinge upon whether useful information is conveyed when an agent sells reserves and then experiences a stockout. If this event implies that it is highly probable that the observed agent is of the bad type, then the expected value of firing the agent and drawing a new one from the pool will be positive. This expected value must be sufficiently large that it outweighs the expected cost to the principal of distorting the agent's behavior when selling reserves—the threat of being fired will cause the agent to exercise "excess caution" and target a higher reserve level when selling reserves than when buying reserves. Further, we also allow for a fixed cost of firing, C.

The numerical example below shows that, with appropriate choices for parameter values, the good agent will never sell reserves below the maximum possible demand, which is equal to either $\mu_1 + r$ or $\mu_2 + r$, depending on the known distribution of N. However, the bad agent, who does not know which distribution is correct, will find it value-maximizing to hold fewer reserves than the maximum possible demand of $\mu_2 + r$. He thereby accepts a small risk of being fired so that he may avoid large penalties for holding excess reserves. The bad agent will therefore, on occasion, sell reserves below the maximum possible demand, and may experience a shortage after having sold. This difference in behavior allows the principal to perfectly separate the two types. We show that, given this separation, firing a bad agent and replacing him with a new agent from the pool can increase the principal's expected payoff. We also show that the principal will not want to fire an agent that is observed with an insufficient level of reserves following a purchase, as such behavior does not send a clear signal of type.

Setup of numerical example

We will use a specific example with the following values:

$\mu_1 = 9$, $\mu_2 = 10$, and $r = 2$ (these together imply that $N \sim U[7, 11]$ half the time and $N \sim U[8, 12]$ the other half)

$q = \frac{1}{2}$ (the probabilities of drawing either a good or bad agent from the pool are equal)

$s = \frac{1}{2}$ (the bad agent's signal is useless)

R_0 has a discrete distribution and is equal to 8 with probability $\frac{1}{2}$ and equal to 14 with probability $\frac{1}{2}$.

$\tau = 0.05$

$F(R_2 - N) = 2(R_2 - N)$ (linear cost of holding excess reserves)

$G(N - R_2) = 48(N - R_2)$ (linear cost of being short)

$\alpha = \frac{1}{2}$ (half of the principal's payoff is passed through to the agent)

$W = 4$

The optimal choice of R^*

We begin by determining the optimal choice of R^*, from the principal's point of view, under the information sets of both the good agent and the bad agent. That is, we calculate what the principal would direct the agent to procure, had he the ability to do so. Suppose first that it is known that $N \sim U[8, 12]$. The principal's choice of R^* is the solution to the following maximization problem:

$$\max_{R_2} \int_8^{12} \frac{1}{4}[-2(R_2 - N) \cdot I_{R_2>N} - 48(N - R_2) \cdot I_{R_2<N}]dN$$

Here, the $I_{R_2>N}$ and $I_{R_2<N}$ are indicator functions for $R_2 > N$ and $R_2 < N$, and the 1/4 is present because this is the value of the uniform pdf of N on its support interval [8, 12]. This maximization is equivalent to:

$$\max_{R_2} \int_8^{R_2} -2(R_2 - N)dN - 48 \int_{R_2}^{12} (N - R_2)dN$$

Taking the first order condition via Leibniz's rule yields:

$$\int_8^{R^*} -dN + 24 \int_{R^*}^{12} dN = 0$$

The solution to this is $R^* = 11.84$. That is, if the principal knew that $N \sim U[8, 12]$ and chose reserve levels himself, he would choose to hold a reserve level equal to 11.84. While the set-up of this model does not permit the principal to choose reserves directly, note that a variant of the mechanism above, without the fixed wage that can be lost upon being fired, provides the agent with incentives that perfectly reflect the principal's payoffs. That is, the good type of agent, when confronted with such a mechanism, will also choose to hold a reserve level equal to 11.84 when $N \sim U[8, 12]$. In a similar fashion, the good type of agent will hold $R^* = 10.84$ when $N \sim U[7, 11]$.

Now consider the choice of R^* that would maximize the principal's payoffs if the information available were that of the bad agent. In this case, the principal must choose R^* to maximize his payoff given that N may be distributed on either of two intervals. This maximization problem is given by:

$$\max_{R_2} \left\{ \begin{array}{c} \int_7^{11} \frac{1}{8}[-2(R_2 - N) \cdot I_{R_2>N} - 48(N - R_2) \cdot I_{R_2<N}]dN \\ + \int_8^{12} \frac{1}{8}[-2(R_2 - N) \cdot I_{R_2>N} - 48(N - R_2) \cdot I_{R_2<N}]dN \end{array} \right\}$$

Given an initial guess that $R^* > 11$, this is equivalent to the following, in which $\int_7^{11} -48(N - R_2) \cdot I_{R_2<N} dN$ equals zero:

$$\max_{R_2} \int_7^{11} -2(R_2 - N)dN - \int_8^{R_2} 2(R_2 - N)dN - 48 \int_{R_2}^{12} (N - R_2)dN$$

Taking the FOC via Leibniz's rule as before yields that $R^* = 11.68$. That is, if the principal had only the bad agent's information set, he would like to direct the agent to hold a reserve level equal to 11.68. If the principal were to offer a bad agent a payoff mechanism similar to that above, but without the threat of being fired, the bad agent would also choose this level of reserves.

Solution to the example when the agents are sellers

We now turn our attention to the target levels of reserves the agents will choose given the mechanism above (including the threat of being fired). We first focus on the case in which $R_0 = 14$. In this case, the initial level of reserves is higher than the maximum possible level of need under either distribution of N. Therefore, both types of agents will be sellers of reserves. We aim to show two results: first, that the good agent will target a reserve level equal to the maximum possible demand, given his signal, and, second, that the bad agent will target a reserve level below the maximum possible demand given his noisy signal, namely 12.

The good agent, who knows the true distribution of N, will choose a level of reserves that will balance the risk of being short (and therefore fired) against the cost of holding excess reserves. Suppose that the agent receives a signal that $N \sim U[8, 12]$. The maximization problem that he solves is the following:

$$\max_{R_2} \int_8^{12} \frac{1}{4}[-(R_2 - N) \cdot I_{R_2>N} - [24(N - R_2) + 4] \cdot I_{R_2<N}]dN$$

Solving this using Leibniz's rule yields that $R^* = 12$—the maximum level of demand. The cost of being fired is sufficiently high that the risk incurred by holding fewer reserves than the maximum level perfectly balances the cost of holding excess inventory (for a fixed wage lower than $W = 4$, the good agent will hold fewer than 12 units of reserves). Similarly, when the signal is $N \sim U[7, 11]$, the agent will sell reserves until reaching $R^* = 11$. These results imply that the principal will never observe the good agent sell reserves and then fail to meet realized demand.

The bad agent solves:

$$\max_{R_2} \left\{ \begin{array}{c} \int_7^{11} \frac{1}{4}[-(R_2 - N) \cdot I_{R_2>N} - [24(N - R_2) + 4] \cdot I_{R_2<N}]dN \\ + \int_8^{12} \frac{1}{4}[-(R_2 - N) \cdot I_{R_2>N} - [24(N - R_2) + 4] \cdot I_{R_2<N}]dN \end{array} \right\}$$

The solution is $R^* = 11.84$. This implies that, if the $N \sim U[8, 12]$ state of the world is realized, then it is possible for the realized value of N to be high enough (namely, greater than 11.84) that the bad agent is short in Period 3, and therefore fired. As an aside, note that this result holds even when the bad agent's signal is somewhat, though not fully informative—the bad agent may be short for $s \in (0.5, 1)$.

Thus, the good agent will never expose himself to being fired by selling reserves that he might need, but the bad agent will. Therefore, whenever the principal observes that the agent sells reserves but stocks out, he has a perfect signal that the agent is of the bad type. Furthermore, the firing incentive has caused both types of agents to sell fewer reserves than what would be necessary to reach the optimal reserve level from the point of view of the principal (equal to 11.84 or 10.84 in the case of a known high or low signal, and 11.65 in the case of no signal).

Solution to the example when the agents are buyers

The second case to consider is that of $R_0 = 8$, which implies that the agents must buy reserves in order to meet demand. When the agents purchase reserves, the firing penalty does not apply, so they will target reserve levels equal to those that are optimal for the principal. That is, the good agents will target $R^* = 10.84$ or $R^* = 11.84$, depending on their signal, and the bad agents will target $R^* = 11.68$. However, there is a risk that the agents will not be able to achieve their target: their probability of success in purchasing any $y = R_2 - R_0$ is given by $e^{-0.05y}$. Both types of agent may therefore be short of reserves when $R_0 = 8$. When the principal observes the agent purchase reserves, then come up short, he must therefore use Bayes' rule to determine the likelihood that the agent of the bad type. The analysis below shows that this likelihood is not much greater than the underlying probability of drawing a bad type from the pool, equal to $q = 0.5$.

We first determine the probability that a good agent will purchase reserves, and still be short. Taking the case of $N \sim U[8, 12]$, the probability that the agent will successfully procure $R^* = 11.84$ is given by $e^{-0.05(11.84-8)} = 0.8253$. The probability density function of R_2 is therefore given by:

$$f(R_2) = \begin{cases} 0.05\exp(-0.05(R_2 - 8)) & \text{if } R_2 \in [8, 11.84) \\ 0.8253 & \text{if } R_2 = 11.84 \end{cases}$$

Given that $N \sim U[8, 12]$, the probability that the agent will be short conditional on a given R_2 is the probability that the realized value of N will be less than R_2, given by $3 - \frac{R_2}{4}$. The unconditional probability that the good agent will be short is then determined by the following integral, which integrates the conditional probability over the pdf of R_2:

$$\Pr(R_2 < N | goodagent) = \int_8^{11.84} (3 - \frac{R_2}{4}) \cdot 0.05 e^{-0.05(R_2-8)} dR_2 + 0.8253(3 - \frac{11.84}{4})$$

This evaluates to a probability of 0.1265. When $N \sim U[7, 11]$, the good agent targets $R^* = 10.84$, and the probability of a stockout is 0.0881. Thus, a good agent will be short of reserves following a purchase with an average probability of 0.1073.

The bad agent always targets $R^* = 11.68$ when buying reserves. The agent will be able to procure reserves sufficient to reach this level with probability 0.8319, and the overall probability of being short is given by:

$$\begin{aligned} &\frac{1}{2} \int_8^{11.68} (3 - \frac{R_2}{4}) \cdot 0.05 e^{-0.05(R_2-8)} dR_2 + 0.8319(3 - \frac{11.68}{4}) \\ &+ \frac{1}{2} \int_8^{11} (2.75 - \frac{R_2}{4}) \cdot 0.05 e^{-0.05(R_2-8)} dR_2 \end{aligned}$$

The first integral represents the probability of stockout if $N \sim U[8, 12]$, and the second represents the stockout probability if $N \sim U[7, 11]$. This evaluates to a probability of 0.1399.

With these probabilities in hand, we are now in a position to use Bayes' rule to find the probability that, having observed a stockout following a purchase, the agent is bad. We have:

$$
\begin{aligned}
\Pr(badagent|R_2 < N) &= \frac{\Pr(badagent) \cdot \Pr(R_2 < N|badagent)}{\Pr(R_2 < N)} \\
&= \frac{1/2 \cdot 0.1399}{1/2 \cdot (0.1073 + 0.1399)} \\
&= 0.5659
\end{aligned}
$$

Observing a purchase followed by a stockout causes the principal to believe that the agent is bad with a 56.59% probability—a small increase over his prior of 50%.

Payoffs from the mechanism, and the incentive to fire

With the agents' actions now determined under all possible scenarios, we can calculate the expected payoffs to the principal. As an example of such a calculation, the expected payoff to the principal from having a good agent with a low signal who is a seller of reserves, when the firing mechanism is not in place (so that $R_2 = R^* = 10.84$), is given by:

$$
E[Payoff] = \frac{1}{4}[\int_7^{10.84} -2(10.84 - N)dN - 48\int_{10.84}^{11}(N - 10.84)dN] = -3.84
$$

The case of a good agent with a low signal who is a buyer of reserves requires a double integral, to account for the probability distribution of R_2:

$$
\begin{aligned}
E[Payoff] &= \int_8^{10.84} \frac{1}{4}[\int_7^{R_2} -2(R_2 - N)dN - 48\int_{R_2}^{11}(N - R_2)dN] \\
&\quad *0.05\exp(-0.05(R_2 - 8))dR_2 - 0.8676 * 3.84 \\
&= -6.14
\end{aligned}
$$

Proceeding with integrals such as these for all cases yields the following table of results for the principal's payoffs:

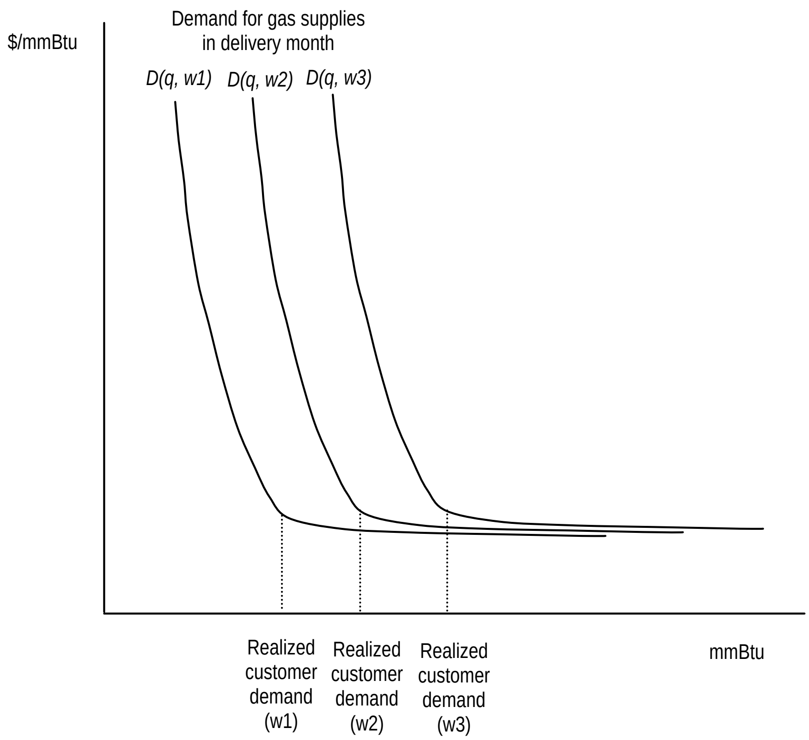
	high signal good agent		low signal good agent		bad agent		good - bad
	R^*	Payoff	R^*	Payoff	R^*	Payoff	
$R_0 = 14$, no firing	11.84	-3.84	10.84	-3.84	11.68	-4.68	0.84
$R_0 = 14$, firing	12.00	-4.00	11.00	-4.00	11.84	-4.76	0.76
$R_0 = 8$ (buying)	11.84	-9.46	10.84	-6.14	11.68	-8.52	0.72
Fire - don't fire		-0.08		-0.08		-0.02	

The first row of the table shows the target R^* for each type of agent, as well as the expected loss to the principal, when the agents are sellers of reserves and the mechanism of transfers from the principal to the agent does not include the threat of being fired. The second row also shows values for the case when the agents are sellers, but indicates behavior when the agents may be fired for selling and then stocking out. The third row illustrates the agents' behavior and the principal's losses when the agents are buyers of reserves (this behavior is invariant with regards to whether the mechanism includes firing or not, since the mechanism calls for firing the agent only if the agent has a stockout after having sold reserves).

The final row provides the expected change in the one-shot payoff to the principal when firing is included in the mechanism. Because the threat of being fired distorts the agents' incentives so that they are excessively cautionary when selling reserves, this mechanism causes a small decrease in payoff. However, the expected payoff of firing a bad agent and replacing him with a new agent from the agent pool must also be considered. The last column of the table indicates the expected difference in payoff between the good agent and the bad agent—the good agent yields substantially greater expected profits to the principal: when the firing mechanism is in place, the additional expected payoff of a good agent is equal to 0.74 (this is the average of 0.72 and 0.76). Given the 50/50 chance of drawing a good agent from the pool, firing the bad agent yields an expected benefit of 0.37, greater than the small distortionary cost of the firing mechanism.

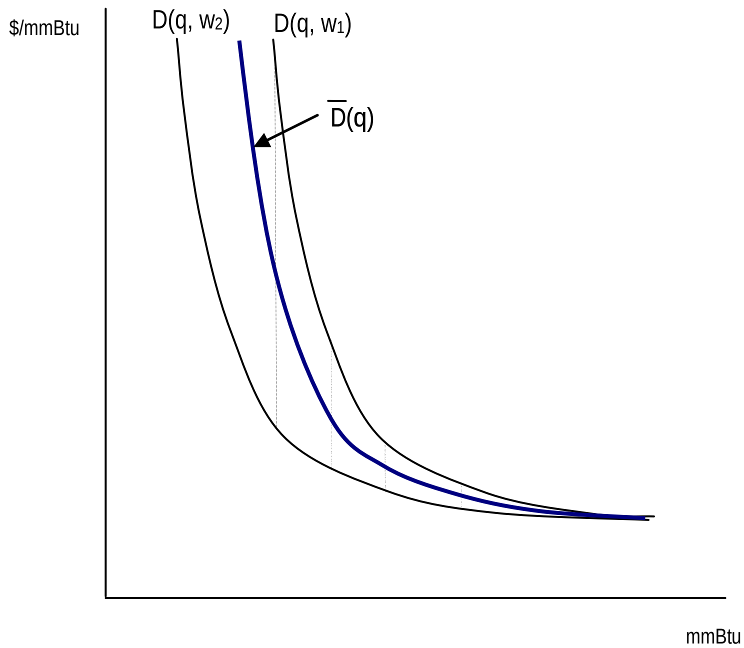
Also note that, given a small fixed cost of firing, the principal will not choose to fire the agent if the agent is observed with a stockout following a purchase of reserves. Recall that such an observation implies that the agent is bad with a probability of only 0.5659. Thus, if the principal fires the agent, there is only a $0.5 \cdot 0.5659 = 0.2830$ probability of actually replacing a bad agent with a good agent, and a $0.5 \cdot 0.4341 = 0.2170$ probability of replacing a good agent with a bad agent! The expected payoff of carrying out the firing is only 0.05. Therefore, as long as the fixed cost of firing, C, is greater than 0.05 but less than 0.37, it is beneficial for the principal to fire the agent when he observes a stockout following a sale, but not when he observes a stockout following a purchase.

Figure 1



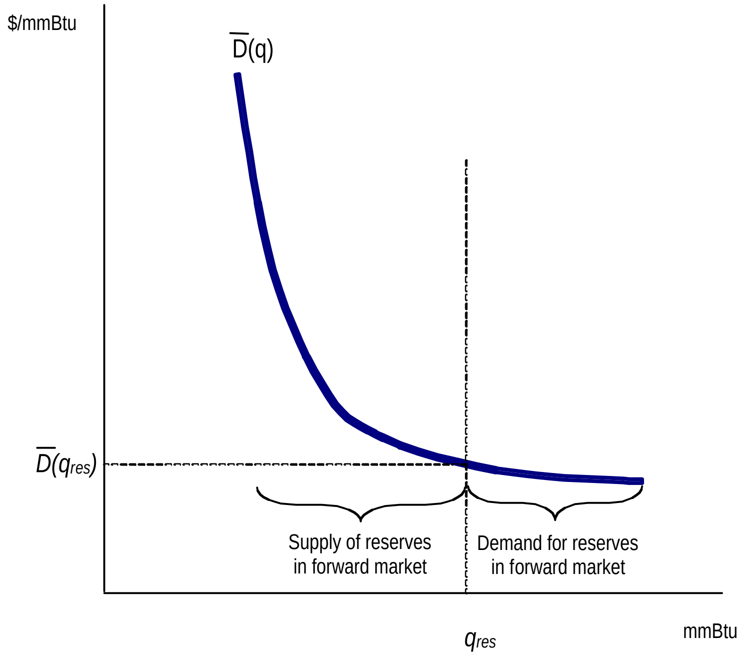
Demand for gas supplies in the delivery month (i.e., once the demand realization is known) is convex, reflecting asymmetric costs of insufficient vs. excess supplies of gas. Realized customer demand and demand curves vary with demand shocks, such as weather.

Figure 2



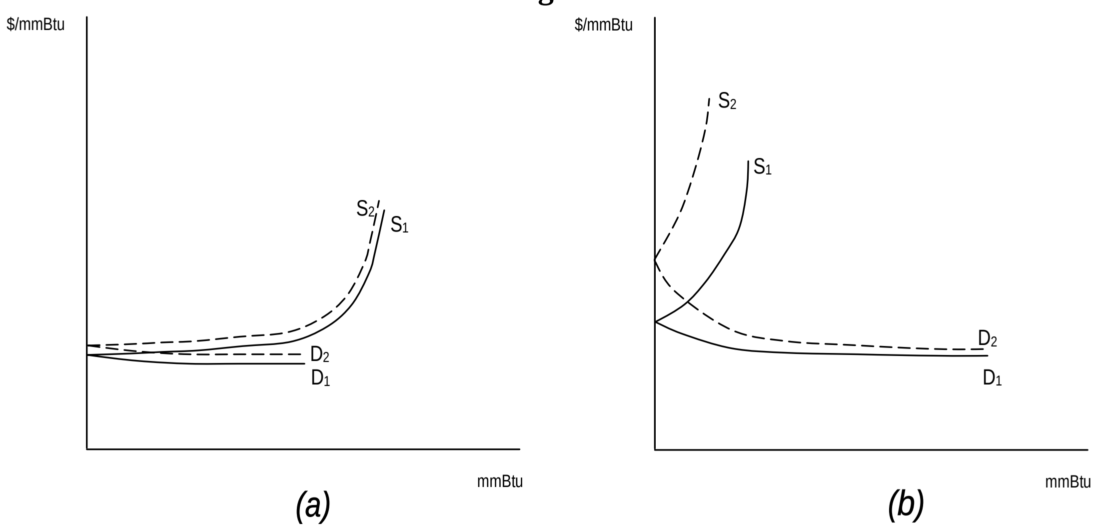
The marginal valuation of any incremental unit of gas supply in the forward market will be the expectation of marginal valuations of that unit of gas under the possible demand realizations. This graph illustrates a hypothetical example with two equally likely demand realizations.

Figure 3



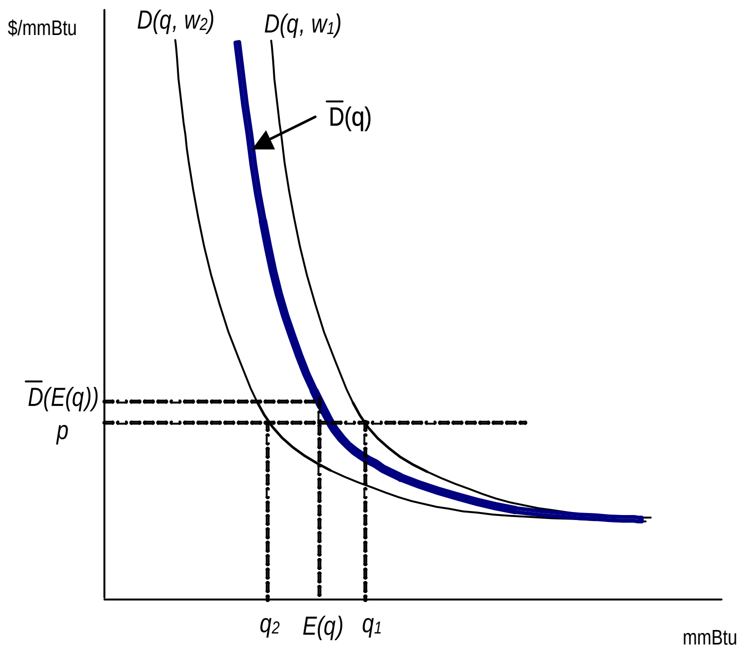
Before the delivery month begins, an LDC will have some quantity of reserves, q_{res}. The marginal valuation of this quantity is $\bar{D}(q_{res})$. If the forward price is above this, the LDC will want to sell reserves according to the marginal valuation schedule, $\bar{D}(q)$, which lies above $\bar{D}(q_{res})$. Conversely, if the forward price is below $\bar{D}(q_{res})$, the LDC will want to buy reserves according to the lower part of the marginal valuation schedule.

Figure 4



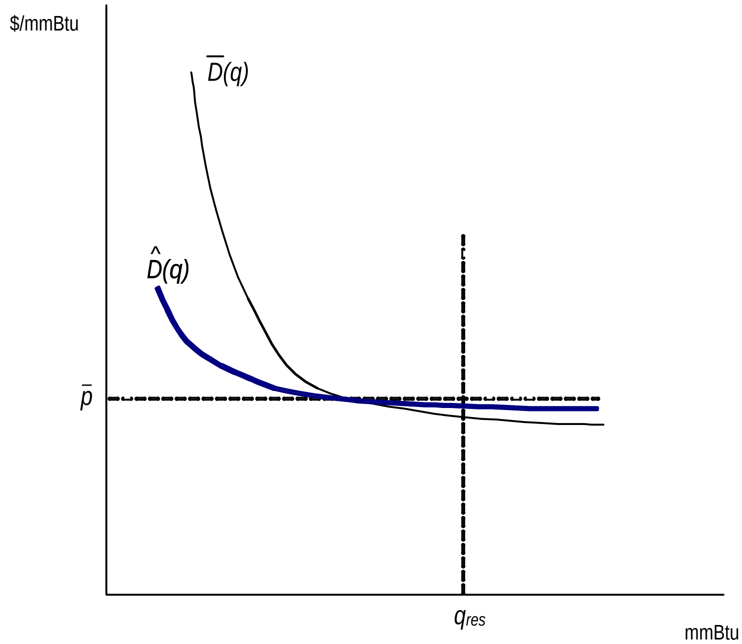
Each panel shows the supply and demand curves in the forward market for two different LDCs. In panel (a), both LDCs have a relatively high level of reserves, meaning that for these firms, q_{res} would lie on the lower, righthand portion of the marginal valuation curve depicted in Figure 3. In panel (b), both LDCs have a relatively low level of reserves, meaning that for these firms, q_{res} would lie on the upper lefthand portion of the marginal valuation curve depicted in Figure 3. Note that in panel (a), the marginal valuation at the vertical axis (which represents the current level of reserves) is very similar for the two firms, while in panel (b), the marginal valuation at the current level of reserves is very different for the two firms.

Figure 5



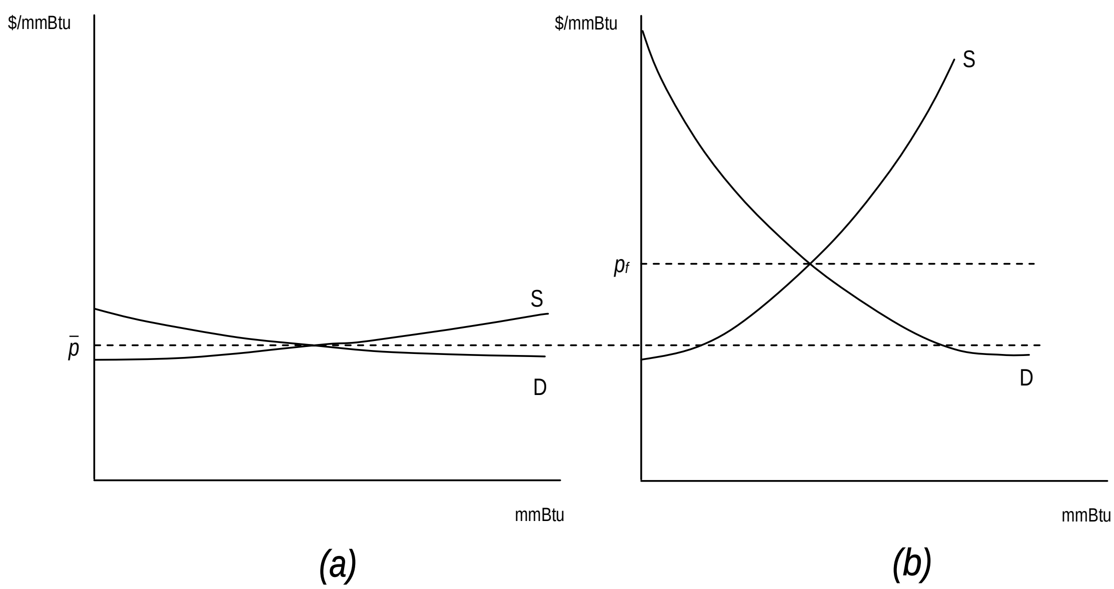
If an LDC faced a price of *p* in the spot market under the high demand condition (w1), it would wish to have supplies of q_1. If it faced the same price of *p* in the low demand condition (w2), it would wish to have supplies of q_2. The expected level of supplies it would wish to have if it had access to a reliable spot market at a spot price of *p* is therefore *E(q)*, the average of q_1 and q_2. In the forward market, the LDC's willingness-to-pay for the quantity *E(q)* is $\bar{D}(E(q))$, which is greater than *p*.

Figure 6



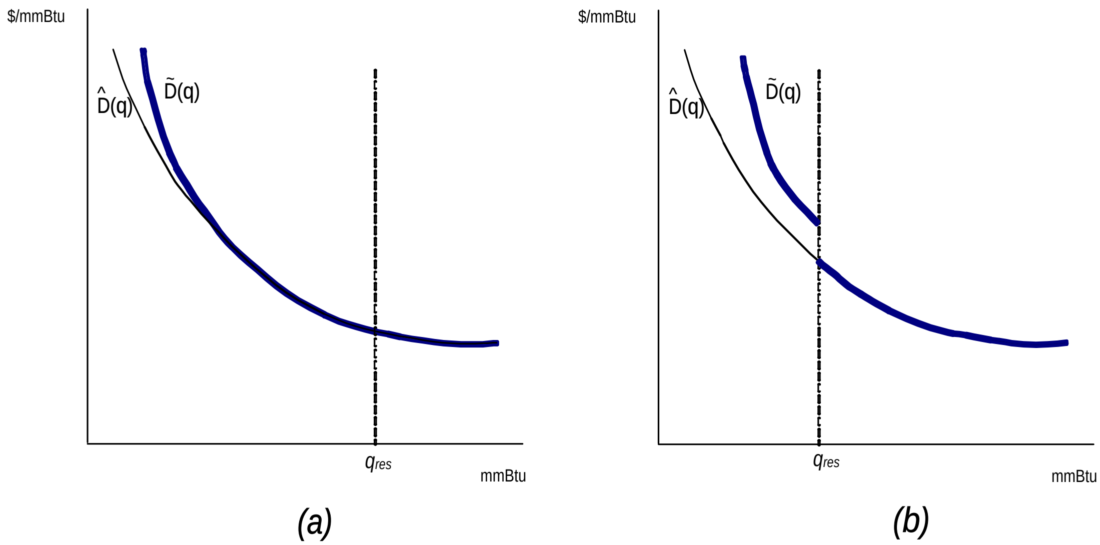
For an LDC that may have access to the spot market, its willingness-to-pay in the forward market is $\hat{D}(q)$, which is the weighted average of $\bar{D}(q)$—the LDC's marginal valuation for gas in the forward market when it cannot access the spot market—and $\bar{p}$, the expected spot price. The weight used to average is the probability that the firm will have access to the spot market. Note that the current level of reserves may fall anywhere along these schedules.

Figure 7



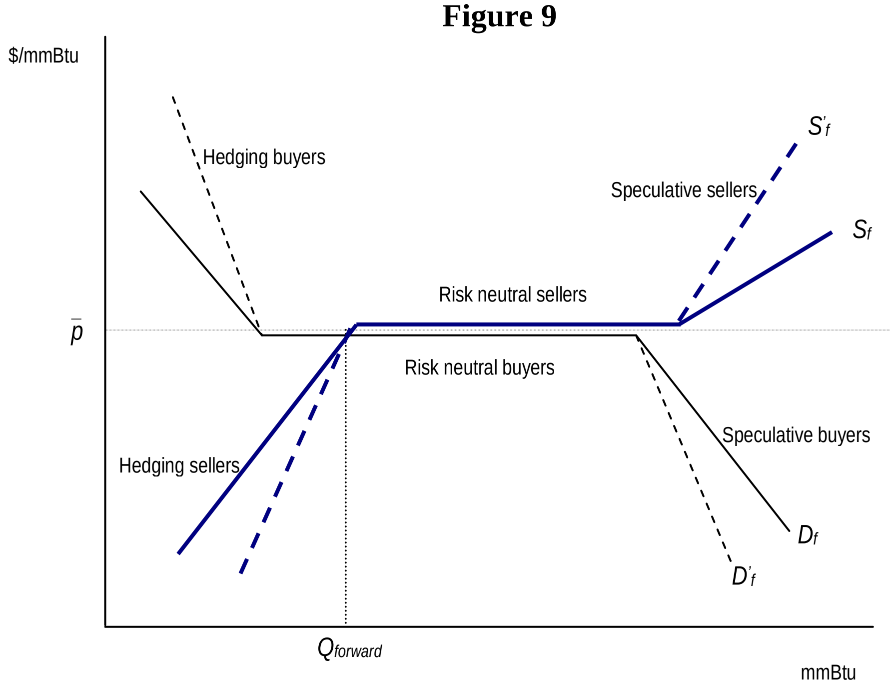
Panel (a) shows the market demand and supply of natural gas (aggregated over all LDCs in the market) in the forward market when the spot market is expected to be loose. Panel (b) shows the forward market supply and demand in when the spot market is expected to be tight.

Figure 8



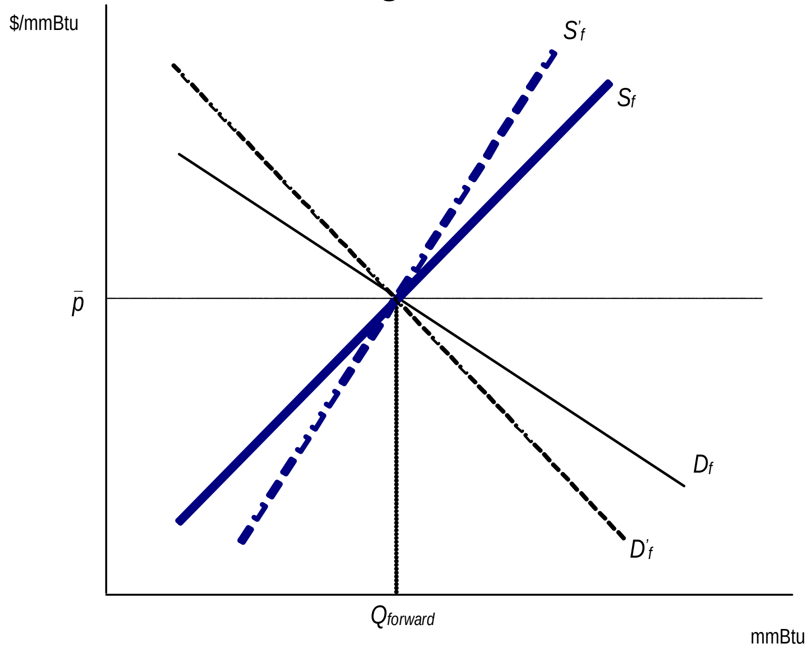
In both panels, $\hat{D}(q)$ represents an LDC's willingness to buy and sell in the forward market if there is no extra regulatory penalty associated with having to curtail customers *after* having sold reserves in the forward market and then being unable to procure them in the spot market. $\tilde{D}(q)$ represents the willingness to buy and sell if an LDC does face such a regulatory penalty. Panel (a) shows an LDC with abundant reserves relative to the range of realized customer demands it thinks likely. The LDC in panel (a) could sell reserves in the forward market with a near zero increase in the probability of incurring the regulatory penalty. Panel (b) shows an LDC in a tight reserve position. For the LDC in panel (b) any sale of reserves in the forward market will result in a non-negligible increase in the probability of incurring the regulatory penalty.

Figure 9



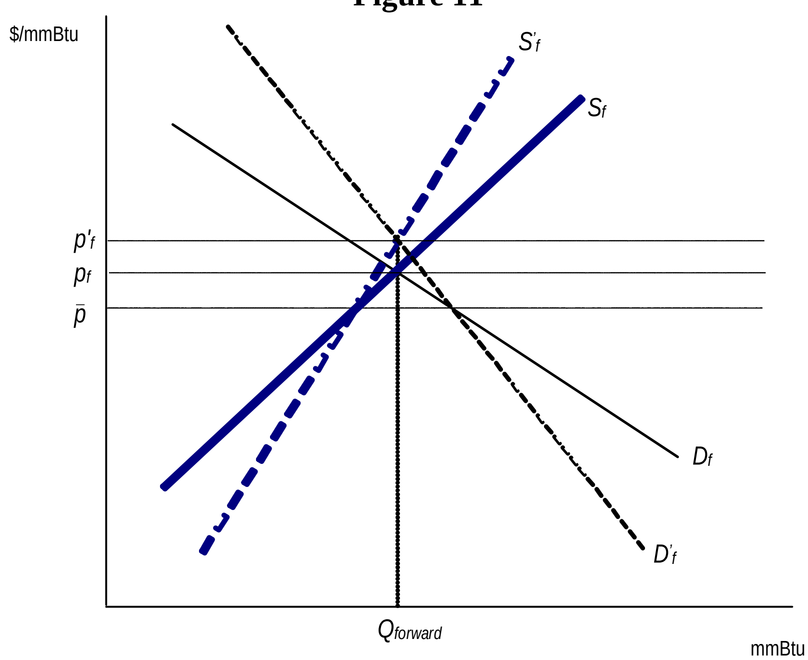
A hypothetical forward market with a large number of risk neutral, and some risk averse, buyers and sellers. An increase in the variance of spot prices is represented by the shift from D_f to D'_f and from S_f to S'_f. The effects of risk aversion are exacerbated: hedging buyers are willing to pay more and speculative buyers willing to pay less, while hedging sellers are willing to sell for lower prices and speculative sellers hold out for higher prices.

Figure 10



A hypothetical forward market with no risk neutral traders. This graph shows a special case in which the forward price equals the expected spot price. In this case, increased price volatility in the spot market has no effect on the forward price or on the volume of trade in the forward market.

Figure 11



A hypothetical forward market with no risk neutral traders. This graph shows a case in which there is a forward price premium. Increased volatility in the spot price would increase the forward price premium, but does not change the volume of trade in the forward market.

Table 1: Spot and Bidweek Data Summary Statistics

	Number of observations	Number of locations	Median	Mean	Std Dev	Min	Max
Spot Price ($/mmbtu)	11619	131	2.71	3.68	2.26	0.87	23.96
Bidweek Price ($/mmbtu)	11619	131	2.69	3.67	2.30	0.82	19.76
Henry Hub Future Price ($/mmbtu)	174	n/a	2.92	4.04	2.47	1.42	14.27
Basis Differential to Henry Hub ($/mmbtu)	11619	131	-0.08	-0.13	0.72	-2.16	26.11
Bidweek Volume (million cubic ft / month)	5062	108	245	465	562	1	4992

Table 2: Full Sample and Rolling Regression Results for Determinants of Spot Price

	Full Sample Results†	Mean Coef Over Valid RR Predictions (174 regressions)	Std Deviation of Coef Over Valid RR Predictions (174 regressions)
HH future price	0.8738 (0.0564)	0.8589	0.0737
Node to HH basis differential	0.3633 (0.1232)	0.2845	0.1368
N	17438	-	-
R^2	0.8688	-	-

Regressions include location fixed effects

†Standard errors are clustered on year-month

Table 3: Determinants of Bidweek Price Minus Spot Price

	1 IV	2 IV	3 IV	4 IV, GLS	5 IV, GLS	6 IV, GLS
Spot price (instrumented with prediction)	0.2120 (0.1259)	0.1982 (0.1070)	0.2029 (0.1090)	0.1871 (0.1116)	0.1894 (0.1026)	0.1940 (0.1047)
Location fixed effects	Y	Y	Y	Y	Y	Y
4th order polynomial in year-month	Y	Y	Y	Y	Y	Y
Month-of-year fixed effects	N	Y	Y	N	Y	Y
Linear time trend * location interaction	N	N	Y	N	N	Y
ρ	-	-	-	-0.064	-0.024	-0.023
N	11619	11619	11619	11619	11619	11619

GLS regressions use Prais-Winsten

Standard errors are clustered on year-month

Table 4: Determinants of Bidweek Volume

	1 IV	2 IV	3 IV	4 IV, GLS	5 IV, GLS	6 IV, GLS
Spot price (instrumented with prediction)	-0.0789 (0.0135)	-0.0761 (0.0128)	-0.0622 (0.0116)	-0.0986 (0.0451)	-0.0771 (0.0299)	-0.0693 (0.0175)
Location fixed effects	Y	Y	Y	Y	Y	Y
4th order polynomial in year-month	Y	Y	Y	Y	Y	Y
Month-of-year fixed effects	N	Y	Y	N	Y	Y
Linear time trend * location interaction	N	N	Y	N	N	Y
ρ	-	-	-	0.535	0.540	0.342
N	5062	5062	5062	5062	5062	5062

GLS regressions use Prais-Winsten

Standard errors are clustered on year-month

Figure 12

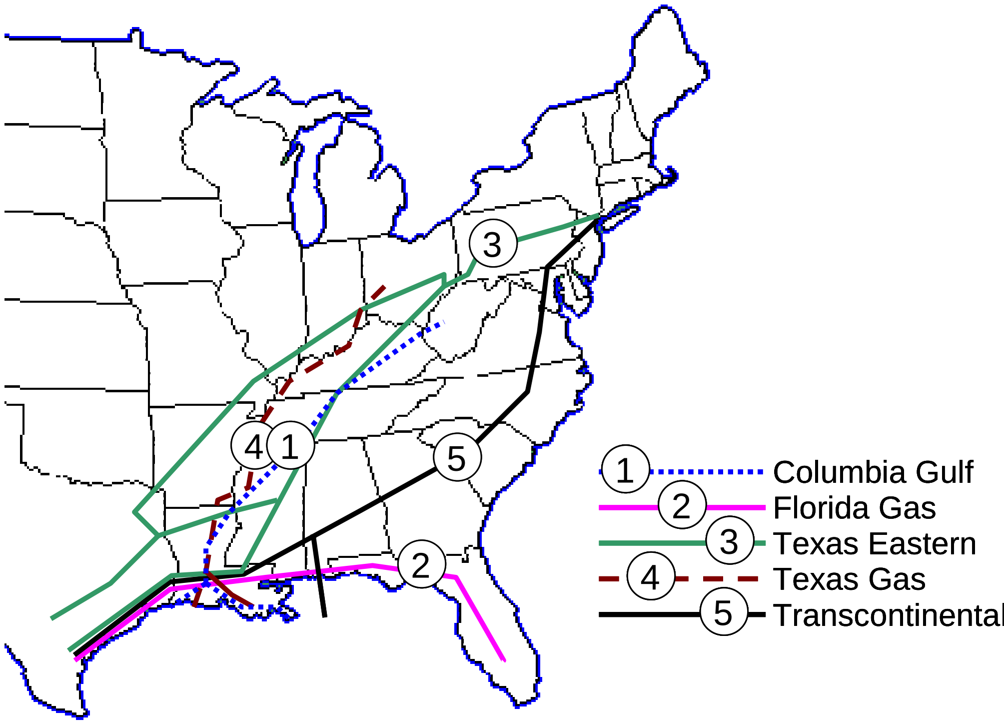


Table 5: Number of Products and Capacity Releases by Pipe

Pipe Number	Pipeline	# of Routes	# of Capacity Release Products	# of Non-Recallable Capacity Releases	# of Recallable Capacity Releases
1	Coumbia Gulf	2	28	72	64
2	Florida Gas	3	30	11	50
3	Texas Eastern	4	51	36	168
4	Texas Gas	1	18	16	129
5	Transcontinental	2	28	103	45
Total		12	155	238	456

Table 6: Determinants of Number of Non-Recallable Capacity Releases Per Month
Poisson Regressions

	1		2		3	
	Incidence Rate Ratio	p-value	Incidence Rate Ratio	p-value	Incidence Rate Ratio	p-value
Spot margin	0.1149	0.053	0.1157	0.060	0.2436	0.155
Route fixed effects	Y		Y		Y	
Linear time trend	N		Y		Y	
Month-of-year fixed effects	N		N		Y	
N	155		155		155	
Log likelihood	-199.7		-198.9		-184.2	

Standard errors are clustered on year-month

Table 7: Determinants of Number of Recallable Capacity Releases Per Month
Poisson Regressions

	1		2		3	
	Incidence Rate Ratio	p-value	Incidence Rate Ratio	p-value	Incidence Rate Ratio	p-value
Spot margin	0.4666	0.274	0.4677	0.278	0.4516	0.195
Route fixed effects	Y		Y		Y	
Linear time trend	N		Y		Y	
Month-of-year fixed effects	N		N		Y	
N	155		155		155	
Log likelihood	-310.5		-310.4		-299.3	

Standard errors are clustered on year-month