The forward premium in electricity markets: an experimental study

Silvester van Koten**

FSE-UJEP, University J. E. Purkyně, Moskevská 54, 400 96 Ústí nad Labem

CERGE-EI, a joint workplace of Charles University and the Economics Institute of the Czech Academy of Sciences, Politickych veznu 7, 111 21 Prague, Czech Republic.

Abstract

An economics laboratory experiment is used to test the validity of Bessembinder and Lemmon's 2002 seminal risk premium theory. The theory predicts that forward premia in electricity markets are determined by the statistical properties of demand. The existing empirical evidence is mixed, possibly as a result of the lack of observability of key variables. Specifically, the experiment tests if an increase in the variance of demand makes the forward premia more negative for specific parameters and implementation details. The experimental results corroborate the theoretical predictions.

Keywords: Forward Premia, Electricity Markets, Economics Experiments

JEL codes: C92, G13, G40, L94, Q47

1. Introduction[1]

Bessembinder and Lemmon's 2002 (hereafter referred to as BL) risk premium theory links present forward electricity prices to the statistical properties of anticipated electricity demand. The theory and its predictions are important for participants in electricity markets, as financial markets are essential for their risk management (see, for example, Bun and Chen, 2013). Outside the financial markets, few other options for risk management of the electricity market exist, as it is near impossible to store electricity in significant quantities. Moreover, the non-storability of electricity, together with the high variability of net electricity demand in the short term, leads to extreme volume and price volatility in the electricity spot markets,[2] thus resulting in very high risks for the market participants.

Driver of the risk is the variation in electricity demand and the particular characteristics of producers and retailers. A particular characteristic of producers in the electricity industry is that they have convex cost functions (e.g., Harris, 2006; BL). The price, and thus their profits, vary strongly with changes in demand. A particular characteristic of retailers is that they are obliged to fulfill the full demand of their customers for a fixed retail price (an uplift on the expected wholesale price) (e.g., BL). The occurrence of a higher demand than expected may strongly increase the wholesale price, resulting in a lower (possibly even negative) profit.

** Senior Researcher at the Jan Evangelista Purkyně University. Email: Silvester.VanKoten@ujep.cz. The paper was written as part of the project financed by UJEP-IGA-TC-2019-45-01-2 for the project "Economic Experiments". Financial support from GACR for the initial parts of the research is gratefully acknowledged. I am grateful to Andreas Ortmann and Sebastian Schwenen for their helpful comments. A special thanks to Jan Vavra for his software support, especially the front-end development of the software in javascript/react. The usual disclaimer applies.

[1] Part of the introduction is based on Van Koten (2020).

[2] The price volatility of electricity can be two orders of magnitude higher than for other commodities or financial instruments (Weron, 2007).

Producers (retailers) thus bear considerable risk and wish to hedge against this risk by selling (buying) forwards, and increasingly so with a higher variation of demand.

The importance of risks for market participants is a relatively new phenomena. In the past, electricity producers and retailers were partly sheltered from risk by the model of vertically integrated utilities operated as regulated (state) monopolies, but in the present liberalized model of unbundling and competition, both electricity producers and retailers must now each shoulder the full price and volatility risks themselves. Further, possibly as a result of the extreme volatility, trading in financial electricity derivatives has increased tremendously over the past 15 years.[3] The role of financial markets for risk management in the electricity market can be expected to gain even further in importance in the near future, as the variability of net electricity demand in the short term has been increasing over time due to an drastic increase in intermittent generation.[4]

BL's theory plays a central role in the modeling of electricity prices in spot and forward markets (Longstaff and Wang (2004), Karakatsani and Bunn (2005), Diko, Lawford and Limpens (2006), Hadsell and Shawky (2006), Douglas and Popova (2008), Lucia and Torro (2008), Weron (2008), Daskalakis and Markellos (2009), Redl, Haas, Huber and Böhm (2009), Botterud, Furio and Meneu (2010), Kristiansen and Ilic (2010; Haugom and Ullrich, (2012), Bun and Chen (2013), Handika and Trück (2013), Redl and Bunn (2013), Weron and Zator (2013), Zator (2013), Fleten et al. (2015), Jacobs and Pirrong, (2017). and Xiao et al. (2015) Jones (2018)). BL's theory guides modeling and is – to an extent – a litmus test for the results of empirical analysis on electricity spot and forward prices (see for example Karakatsani and Bunn (2008) on spot market price formation). However, empirical tests of BL's theory have been equivocal. While some empirical studies provide support, other studies' support is weak or presents outcomes opposite to the theoretical prediction. The equivocal outcomes of empirical tests may be the result of methodological issues that make it difficult to test the theoretical predictions (Weron and Zator, 2014, Zator, 2013). These issues are addressed in more detail in Section 2.2.

An experimental test can therefore provide additional guidance. An economics experiment provides a controlled environment that can be carefully designed to assure that the relevant theoretical assumptions are fulfilled. According to my best knowledge, no earlier attempt has been done to test BL's theory experimentally. The results of the experiment performed in this paper support BL's theory for the specific parameter choices and implementation details. I describe BL's theory and the empirical literature in section 2. I describe the experimental design and test in section 3. I present the results in section 4 and conclude in section 5.

[3] For example, in the EEX market, trading in European electricity derivatives increased thirtyfold from a level of 119 TWh in 2002 to 3,347 TWh in 2018 (EEX, 2005, 2019).

[4] The deployment of intermittent renewable power plants such as photovoltaic solar panels and wind mills has increased the variance of the net demand (the demand minus the production of renewables). See, for example, Staffell and Pfenninger (2018). For example, in the EU, the official EU policy mandates a further dramatic increase in the deployment of intermittent renewable power plants (EU Commission, 2014), thus further increasing the variability of net electricity demand.

2. Bessembinder and Lemmon's 2002 forward premium theory

2.1 Theory

The forward premium is defined as the difference between the forward price (the price today of a unit to be delivered in the future) and the expected future spot price. More precisely:[5]

$$FP_{t_0} = P_{F,t_0,t_1} - E_{t_0}[P_{S,t_1}]. \tag{1}$$

In Equation (1), t_0 refers to the present period, t_1 to the future period, P_{F,t_0,t_1} to the present (time t_0) price of a forward contract with delivery at time t_1, P_{S,t_1} to the future (time t_1) spot price, and $E_{t_0}[\cdot]$ to the present (time t_0) expectation operator for future (time t_1) outcomes. For ease of notation and consistency with BL, I drop the time–indexes below. The relative forward premium can now be defined as:

$$RFP = FP / E[P_S]. \tag{2}$$

Assuming risk-neutrality and ignoring the interest rate, the forward premium would be zero as the forward price is then an unbiased estimator of the future to-be-realized wholesale spot price.[6] With risk-aversion, the forward premium can generally be expected to be different from zero. When retailers bear more (less) risk than producers, they should be more (less) eager to hedge than producers, thus increasing (decreasing) the forward price above (below) the expected spot future price, resulting in a positive (negative) forward premium. BL model this issue by formulating the profit functions of producers and retailers, assuming that production cost functions are strictly convex[7] and that retailers are obliged to fulfill the full demand of their customers for a fixed retail price (an uplift on the expected wholesale price). They further assume that there is no uncertainty in spot markets, that the electricity industry is perfectly competitive, and that the time between markets is short enough to ignore the interest rate.

A further modeling assumption is that all electricity producers have identical convex cost functions given by $C[q] = f + \frac{a}{c} q^c$. The parameter $c > 2$ is the cost convexity parameter; *f* is the fixed cost and the variable cost parameter *a* is a scaling parameter in the cost function of electricity producers. Further the following notation is

---

[5] See Weron and Zator (2014) for alternate definitions and a discussion of the confusion the interchangeable use of different definitions has caused in the literature.

[6] Another well-known equation in forward markets ties the present forward price to the present spot price though arbitrage with storage: $P_F = (P_S + c)e^{(r-y)t}$, where P_S is the spot price, P_F the forward price, *c* the storage cost, *r* the interest rate, *y* the convenience yield, and *t* time. Electricity is, however, not storable, and facilities to transform electricity into a storable form and back (such as batteries or pumped storage) are rare as they are inefficient and expensive. Therefore, this equation is generally not valid in the vast majority of electricity markets.

[7] Producers generally have power plants with widely varying marginal costs among their generating assets as the plants are of different types, use different kind of fuels and are of different vintages. Moreover, even when a given plant is running on full capacity, producers can squeeze out still some more electricity, but at the (very high) cost of a reduced plant lifetime (Harris, 2006, p.51 and p.485-487). The convexity of the cost function is thus determined by the precise composition of the power plant portfolio. Optimization approaches as in Soft (2002, p.33-45, 123-129) and Biggar and Hesamzadeh (2014), balancing the fixed and variable cost of different plant types against the variability of demand, can give an account of the optimal power plant portfolio.

used. N_P (N_R) is the number of identical producers (retailers); RA is the degree of risk aversion[8] of retailers and producers; $x = \frac{1}{c-1} < 1$; and $P_R = r \cdot \mathrm{E}_{t_0}[P_{S,t_1}]$ refers to the regulated or otherwise temporarily fixed retail price of electricity at which retailers can sell electricity to consumers (with $r > 1$). As competition is assumed to be perfect, spot prices are equal to marginal costs and thus determined by $P = C'[Q/N_P] = a \cdot (Q/N_P)^{c-1}$. Using these assumptions, they derive formula for the supply by producers $q_{F,P}$ and the demand by retailers $q_{F,R}$ for forward positions (BL, p.1379):

$$q_{F,P} = \frac{1}{Var[P_S]}\left(\frac{P_F - E[P_S]}{RA} + \frac{1}{a^x}(1-\frac{1}{c})Cov[P_S^{x+1}, P_S]\right) \tag{3}$$

$$q_{F,R} = \frac{1}{Var[P_S]}\left(\frac{P_F - E[P_S]}{RA} + P_R Cov[q_R, P_S] - Cov[P_S q_R, P_S]\right) \tag{4}$$

Solving Equations 3 and 4 for the forward price P_F such that supply and demand are equal and rearranging then results in (BL, p.1379):

$$FP = -\frac{N_P}{\left(N_P + N_R\right)} \frac{RA}{ca^x}\left(c\overline{P}_R \operatorname{Cov}[P_S^x, P_S] - \operatorname{Cov}[P_S^{x+1}, P_S]\right). \tag{5}$$

Given the statistical properties of the demand distribution, Equation (5) can be used to formulate precise predictions for the sign and size of the forward premium (BL, p.1362):

1. Hypothesis 1: The equilibrium forward premium decreases in the anticipated variance of wholesale prices, ceteris paribus.
2. Hypothesis 2: The equilibrium forward premium increases in the anticipated skewness of wholesale prices, ceteris paribus.
3. Hypothesis 3: The equilibrium forward premium is convex, initially decreasing and then increasing, in the variability of power demand, ceteris paribus.
4. Hypothesis 4: The equilibrium forward premium increases in expected power demand, ceteris paribus.

The first two hypotheses are based on approximations of Equation (5) and the last two hypotheses on simulations using Equation (5). For further details, see Appendix A1 for the equations and Appendix A2 for simulations.

2.2 Empirical literature and evidence[9]

Table 1 gives an overview of the numerous papers that have aimed at testing BL's theory. The literature has been equivocal on the effects of variance and skewness in the forward premia. A number of papers, including Longstaff and Wang (2004), Diko, Lawford and Limpens (2006), Hadsell and Shawky (2006), Douglas and

[8] Assuming a mean-variance utility function ($u[x] = E[x] - RA \cdot VAR[x]$) as in Hirshleifer & Subrahmanyam (1993).

[9] This section is a more comprehensive version of a section in Van Koten (2020).

Popova (2008), Fleten et al. (2015), Jacobs and Pirrong, (2017). and Xiao et al. (2015) reports results that support the theory. Other studies, however, do not support the theory by reporting mostly coefficients for the variance and skewness that are insignificant or with signs opposite to the prediction. Lucia and Torro (2008), Redl et al.(2009), Botterud et al.(2010), and Furio and Meneu (2010) find at best only partial support, while Bunn and Chen (2012), Haugom and Ullrich (2012), repeating the tests of Longstaff and Wang (2004) using the most recent data from PJM, Handika and Trück (2013), Redl and Bunn (2013), Weron and Zator (2014), repeating the test of Botterud et al.(2010) for a larger data set and correcting for methodological and specification mistakes, and Jones (2018) find no support. The empirical studies thus show a mix of supportive and contradictory findings.

Table 1: Overview of empirical tests of Bessembinder & Lemmon (2002)

Empirical tests of hypotheses 1 and 2

	Study	Data used
Support	Bessembinder & Lemmon (2002)	Monthly data from the PJM and CALPX markets
	Longstaff and Wang (2004)	Hourly spot and day-ahead prices from PJM
	Diko et al. (2006)	Daily data from EEX, Powernext, and APX
	Hadsell and Shawky (2006)	Day-ahead and real time data from NYISO
	Douglas and Popova (2008)	Day-ahead and real time data from PJM
	Viehmann (2011)	Hourly day-ahead data from EEX and EXAA
	Fleten et al. (2015)	Monthly, quarterly and annual data from Nordic NASDAQ OMX and German/Austian EEX
	Jacobs and Pirrong, C. (2017)	Hourly spot and day-ahead prices from PJM
Partial support	Lucia and Torro (2008),	Weekly contracts from Nord Pool
	Redl et al.(2009)	Monthly contracts from EEX and Nord Pool
	Botterud et al.(2010)	Weekly contracts from Nord Pool
	Furio and Meneu (2010)	Monthly data from the Spanish OMEL market
No support	Bunn and Chen (2013)	Daily and monthly data from the British market
	Haugom and Ullrich (2012)	Day-ahead and real time data from PJM
	Weron and Zator (2014)	Weekly contracts from Nord Pool (Repeating Botterud et al., 2010)
	Redl and Bunn (2013)	Month-ahead futures from EEX
	Ronn and Wimschulte (2009)	Day-ahead and intra-day data from EEX and EXAA
	Handika and Trück (2013)	Quarterly and yearly data from the Australian market
	Jones (2018)	Hourly spot and day-ahead prices in MISO

Empirical tests of hypothesis 4

	Study	Data used
Support	Bessembinder & Lemmon (2002)	Monthly data from the PJM and CALPX markets
	Karakatsani and Bunn (2005)	Day-ahead and intra-day data from the British market
	Lucia and Torro (2011)	Weekly contracts from Nord Pool
	Furio and Meneu (2009)	Monthly data from the Spanish OMEL market
	Handika and Trück (2013)	Quarterly and yearly data from the Australian market
	Xiao, Colwell, and Bhar (2015)	Daily and 2-monthly data from the PJM market

These contradictory findings may be the result of the methodological difficulties in empirically testing BL's theory as its predictions contains several unobservable elements which may obscure or bias empirical estimates (see, for example, Bun and Chen, 2012; Weron and Zator, 2014 and Zator, 2013). Most importantly, to determine the forward premium, the forward price and the ex-ante expected spot price must be determined. Empirical analysis can, however, only observe the ex-post realized spot price, which may well be different from the ex-ante expected spot price (Bun and Chen, 2013). Empirical studies thus generally measure the ex-post realized spot price and assume that market participants are right on average. In general, however, this will cause the observed ex-post realized spot price to be correlated with the error and estimated effects will be biased (Weron and Zator, 2014).

Some studies try to model the ex-ante expected spot price or to incorporate the main “nuisance variables” that may explain the difference between the ex-ante expected and the ex-post realized spot price (such as reserve margins, water levels in hydro plants, unplanned outages of large power plants), but it is unknown to what degree this empirical strategy solves the problem of the unobservability of the ex-ante expected spot price (Weron and Zator, 2014). Due to these methodological difficulties, empirical studies may thus be less reliable, which may explain why the empirical evidence for BL's theory is mixed.

3. Experimental tests

3.1 Introduction

Economics experiments are well-suited to control the main variables and exclude “nuisance variables”. Using experiments to test BL's theory, the experimenter can specify the demand distributions used for the experimental treatments and thus influence the ex-ante spot price expectations.

Using economics experiments also has drawbacks. Compared to using empirical studies, it is usually more expensive and labor-intensive to generate data. This drawback is especially relevant for testing BL, as a long session with many subjects results in only a few independent observations. Also, it is not a simple task for student subjects, unfamiliar and untrained in trading, in their roles as producers or retailers, to fully appreciate the intricacies of risk emanating from a particular demand distribution and then to correctly implement their insights in a trading strategy. It is thus not impossible that a task may be too difficult for subjects.[10] Therefore, for this experiment, a careful preparation was organized to assure that the participating subjects have a detailed understanding of the task.[11]

At the heart of BL's theory is Equation (5), specifying the relationship between the forward premium and the mean and variance of demand. The focus will be especially on Hypothesis 3, the prediction that the variance of demand affects the forward premium.[12] This hypothesis seems also particularly relevant as the increasing share of intermittent generation (especially in the form of wind and solar generators) is increasing the variance of net demand (see, for example, Staffell and Pfenninger, 2018).

[10] Many examples can be found in the domain of auction theory. Subjects' bidding behaviour in auctions is often not in line with the theoretical predictions (see for example Kagel and Levin, 2015).

[11] The preparation involves a careful design and presentation of the full and summary instructions, an instructive simulation, test questions with feedback, and selection of the 2/3 best scoring subjects on the test. See section 3.3 for further details.

[12] While testing Hypothesis 4, the prediction that the equilibrium forward premium should increase in expected power demand, would also be a viable choice, this was judged less topical than Hypothesis 3. Also, Van Koten (2020) shows that Hypothesis 4 is not generally true: See Appendix A2 for a summary. Hypothesis 1 and 2, regarding the effects of the variance and skewness of wholesale prices on the forward premium, are approximations of the central Equation 5 and testing these hypotheses would be a more indirect test of the core theory. Moreover, the skewness and variance of prices cannot easily be manipulated. The mean and variance of demand can be manipulated, but to use that to manipulate just one of the properties of prices (e.g. the skewness), while keeping the other one constant (e.g. the variance) is not straightforward.

Hypothesis 3: The equilibrium forward premium is convex, initially decreasing and then increasing, in the variability of power demand, ceteris paribus.

3.2 Experimental design and hypotheses

3.2.1 Trading environment and general parameters

To be an effective and internally valid experimental test, the settings and parameters must meet the theoretical assumptions of BL's theory. Ideally, they should also reflect the empirical regularities found in the power markets as much as possible. The main basic parameters that are held constant across all treatments and their chosen values are shown in Table 2.

Table 2: Basic parameter values (all treatments)

Trading environment	Continuous double auction market (CDA)
The number of producers	4
The number of retailers	4
Type of demand distribution	Uniform distribution with mean 60
Cost convexity parameter	4
Cost scaling parameter *a*	0.018
Fixed costs *F*	0

The trading environment used is the continuous double auction market (CDA) which is a very competitive trading environment, as required by the theoretical assumptions. Smith et al. (1982), McCabe et al. (1993), Friedman and Ostroy (1995) and Cason and Freidman (2008), show that the CDA is the trading institution that produces prices and allocations the nearest to the competitive equilibrium, even with relatively few buyers and sellers. The CDA is also an ubiquitous trading environment used in many exchange markets, including power markets such as intraday markets and power future markets.[13]

The number of producers and retailers is chosen to be four of each. In one trading group are thus eight subjects in total. While the theory assumes perfect competition,[14] the number of participants in one market is limited by practical and research funding considerations, especially as each market will result in just one independent observation. However, a CDA with four producers and four retailers can be expected to leads to very competitive outcomes (Smith et al., 1982; McCabe et al., 1993; Friedman and Ostroy, 1995; Cason and Freidman, 2008). In addition, to prevent producers from exerting possible market power, producers are given a production requirement that mostly mirrors the obligation that retailers have to satisfy the full retail demand.

[13] Some power markets use different trading environments. For example, day-ahead markets are often organized as call auctions (sometimes also called clearinghouse auction (Friedman and Ostroy, 1995)), while intraday markets trading or energy futures trading may also take place bilaterally (OTC trade). As this is a first attempt to establish, in an experimental setting, a market premia effect per se, the possible effects of different trading environments on markets premia is left as a future research direction.

[14] The assumption of perfect competition is not always fulfilled in electricity markets, and it could be argued that a HHI between 2000 and 3000 better reflects the competitiveness of electricity markets around the world (see Appendix A4).

See section 3.4 under "Trading in the double auction" for further details.

The distribution of market demand will be given by an uniform distribution with a mean of 60. As there are four producers and four retailer, it means that each market participant faces an uniform distribution with a mean of 15 (60 divided by 4). While, in power markets, the distribution of market demand may take more complicated forms (see, for example Hyndman & Fan, 2010), the theoretical assumptions do not require a specific distribution. I therefore judged that the simplest possible distribution, the uniform one, best serves the objective of internal validity, requiring that experimental subjects understand the effect of the demand distribution on outcomes and prices.[15] To focus on the effect of the standard deviation, the mean demand is held constant (set at 60) and the interval of the uniform distribution is adapted to adjust the standard deviation.

I set the cost convexity parameter equal to four. This parameter reflects the degree of convexity of the cost function, and BL use this parameter value in their simulations for inferring Hypotheses H3 and H4. I set the cost scaling parameter a equal to 0.018, as theory then predicts a market price of 60 when the demand variance is zero.[16] The fixed costs are set to zero to simplify the problem facing subjects in the role of producers. Then the cumulative production costs for the producers can be determined with the function $C[q] = \frac{a}{c} q^c$, and the marginal costs are calculated as $MC[0] = 0$ and $MC[q] = C[q] - C[q-1]$. See Appendix A5 for the specific values of producers in the experiment.

3.2.2 Specific Parameters and numerical predictions

The main specific parameters that differ across the treatments and their chosen values are shown in Table 3, together with the numerical predictions.[17]

Table 3: Specific parameters and predictions

			Treatments		
			T1 (low variance)	T2 (medium variance)	T3 (high variance)
Specific parameters	Interval of the UD		55-65	40-80	20-100
	Retail price (uplift on Price Mean)		73 (20%)	88 (30%)	139 (60%)
Realizations	Spot Market Demand Mean (standard deviation)		60 (2.9)	60 (11.6)	60 (23.1)
Predictions	Spot Price mean (standard deviation)		60.4 (9)	66.7 (36)	86.7 (79)
	Relative Forward Premium (%)	RA=0.005	-0.3%	-8.2%	-65.5%
		RA= 0.05	-3%	-82%	-655%
	Forward Position		14.6	15	16.3

[15] To further support the subjects' understanding of the demand distribution and its outcomes, the subjects, in the experiment before the trading, were presented with an instructive simulation that visualized the specific distribution and its outcomes of demand and prices. See below for details.

[16] As $a = 60 \cdot \left(N_p / \overline{D}\right)^{c-1} = 60 \cdot \left(4/60\right)^{c-1} = 4/225 \approx 0.018$.

[17] The numerical predictions are generated with a simulation. The software used can be inspected or downloaded, installed and run from https://github.com/slvstr1/dAuction2_simulation_public.

For the market demand, I choose the intervals of the uniform distribution (UD) as 55-65 (T1), 40-80 (T2) and 20-100 (T3) for the three respective treatments. These distributions have a constant mean market demand, but are increasing in the standard deviation. The retail price is an uplift over the mean price as in BL. For the distribution with a low standard deviation, treatment T1, I chose an uplift of 20%. For the distributions with higher standard deviations, treatments T2 and T3, I choose higher uplifts (30% and 60%, respectively) to lower the bankruptcy probability for the retailer.

Having specified all parameters, general and specific, I use Equation 5 to generate theoretical predictions for the relative forward premium. The main predictions are the relative forward premia (using Equation 2 and 5) and the forward positions (using Equations 3 and 4). As shown in Table 3, the predictions for the relative forward premia (but not for the forward positions) depend on the value of the risk aversion parameter *RA*. As *RA* is the coefficient on the variance of outcomes as in a mean-variance analysis, it is an absolute measure of risk and it is not known what is a precise numeric value for the particular tasks in these experiments. This hampers the selection of treatments as well as possible ex-ante estimates of statistical power.[18] The predictions are thus computed for an (arguably) extremely low (0.005) and an (arguably) extremely high (0.05) value. It is assumed, as does BL's theory and the empirical studies mentioned, that the *RA* does not vary much over the different risk situations represented in the treatments. In the experiment, the expected earning is kept constant, thus making it more likely that the assumption is innocuous. For a better understanding of this assumption, I report the observed level of absolute risk aversion as implied by the observed forward premia in section 4.[19]

The first treatment (T1: low variance) is designed with a low level of variance to exhibit no effect: a zero forward premium. To keep its characteristics the same as the other treatments, the market demand is still drawn randomly, but from an interval so small that the variance is very low, ranging 55-65. Thus, the forward premium is expected to be not significantly different from zero (-0.3% or -0.03% relative to the forward price, depending on the *RA*). Interestingly, theory predicts that even for such low variance – and thus low risk involved, the forward position is with 14.6 units relatively large, given that the expected demand for a participant is equal to 15.

The second treatment (T2: medium variance) is designed with a level of variance to exhibit the predicted effect, a negative forward premium, if the value of the *RA* is high, but not when it is low. If the *RA* is high, the forward premium is predicted to be -82% relative to the forward price. If the *RA* is low, the forward premium is predicted to be -8.2% relative to the forward price. A negative forward premium can thus be expected to be detected if the *RA* is high, but not when it is low. The predicted forward position is with 15 only slightly larger than in T1. The difference with T1 is so small that is unlikely to be detected in an experimental setup.

The third treatment (T3: high variance) is designed with a level of variance to exhibit the predicted effect, a negative forward premium, regardless whether the value of *RA* is high or low. Indeed, even when the value of

[18] Ex-post power estimates are not calculated as the general consensus is that they are not informative and prone to misinterpretation. See, for example, Hoenig and Heisey (2001).
[19] I am grateful to Sebastian Schwenen for suggesting this analysis.

RA is low, the forward premium is still predicted to be considerable, -65.5% relative to the forward price. The intention is that testing this treatment against a zero effect will be sufficient powered even with a relative high variance and modest sample size. The predicted forward position is 16.5, only slightly larger than in T1 or T2. The difference is so small that it is, again, unlikely to be detected. Overall, the experiment is thus not designed to detect the differences in the forward positions. However, the forward position is predicted to be considerably larger than zero and I thus expected this effect to be significant in the experiment.

While not necessary for the internal validity of the experiment, the predicted values of mean prices and forward premia in the three treatments are within realistic ranges.[20] The mean prices of the three treatments range from 60.4 to 86.7 and are thus in a reasonably realistic range. The intervals of the uniform distribution also present appropriate ranges of variance, ranging from 9 to 79. The predicted forward premia in the experimental treatments range from -0.3% to -65% (for low RA), and are thus also appropriate for negative forward premium.

Notably, no treatments with positive forward premia are generated, as it is complicated to design experiments with positive forward premia while still having retailers with a positive expected profit. For positive forward premia, retailers must bid up the price in the forward market above the expected spot price. Calculations show that retailers do this only under extremely high risks in the spot market, resulting in negative expected profits for most parameter ranges. Dealing with possible negative expected profits would further complicate the experiment. Therefore, in this first experimental test of BL, the focus is exclusively on parameters with positive expected profits for the participants and thus where the theory predicts negative forward premia.

3.2.3 Hypotheses

Applying Equation (5) and using the parameter values as described above, assuming that the RA is relatively stable, an increase in the standard deviation of demand results in a more negative forward premium. The following hypotheses can be formulated:

H1: The forward premia in the treatments will obey:

$$FP_{T_1} = 0$$
$$FP_{T_2} \leq 0$$
$$FP_{T_3} < 0$$

H2: The forward premium in the treatments can be ordered as following:

$$FP_{T_1} > FP_{T_2} > FP_{T_3}$$

[20] Hadika (2012) reports the mean electricity price for Australia to be between A$36 and A$70, but with spikes up to A$4600. Redl and Bunn (2013) average yearly spot price for the years 2004-2009 for some of the main EU markets (APX, EXAA, EEX, Powernext, Belpex, OTE, PolPX) are between EUR 22 and 70. Redl and Bunn (2013) calculates the relative ex-post forward premium for the EEX from 2003 to 2010 and shows that it can range between -55% to +70%. However, the ex-post forward premium, as explained above, is different from the forward-premium, so these values have to be seen as a broad indication of reasonable value ranges for the parameters.

I will test H1 using Wilcoxon one sample signed-rank tests and H2 as pairwise comparisons using Wilcoxon–Mann–Whitney 2-sample rank sum tests.

Also, BL predict higher forward positions in the forward market. However, the predicted differences in forward positions between the treatments are small, and the designed experiment is therefore not likely to detect the differences. Nonetheless, BL predict forward positions considerably higher than zero, and this is the basis for hypothesis H3a. Hypothesis H3b is more precise, and uses the exact numerical values to compare the realized values with the predicted ones. Hierarchical testing will be applied (Laporte et al., 2016), and, for each treatment, hypothesis H3b will be tested only if hypothesis H3a tested significant for that treatment.

H3a: The forward positions in the treatments will obey:

$$FPos_{T_1} > 0$$
$$FPos_{T_2} > 0$$
$$FPos_{T_3} > 0$$

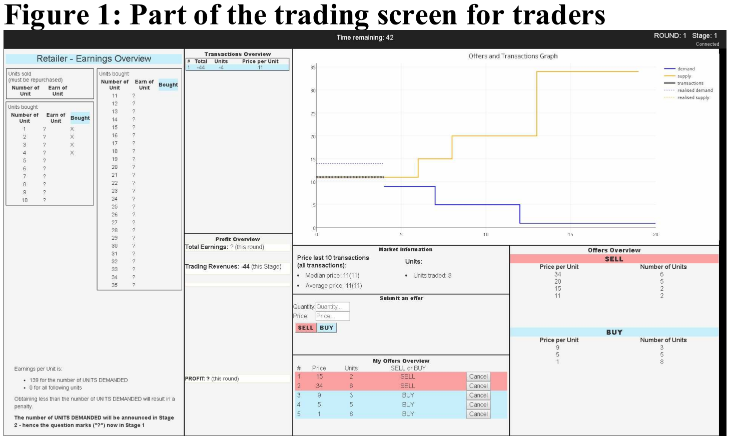
H3b: The forward positions in the treatments will obey:

$$FPos_{T_1} = 14.6$$
$$FPos_{T_2} = 15.0$$
$$FPos_{T_3} = 16.3$$

H3a and H3b will be tested as pairwise comparisons using Wilcoxon–Mann–Whitney 2-sample rank sum tests.

3.3 Experiment Software and procedures

Figure 1: Part of the trading screen for traders



Using Python 3.6/Django 2.08 for the backend and Javascript/React for the front-end, a software product was created that contains, amongst others, a continuous double auction (CDA) with rich visual elements to implement the trading environment.[21] When the CDA is running, the trading book and past transactions are shown in numbers, but are also graphically visualized as shown in Figure 1. The visual elements were added to support the experimental subjects in the process of price discovery.

The software product also contains the full instructions and other elements to prepare and test the experimental participants, such as a training program showing outcomes of random draws of the uniform distribution and the associated market prices under perfect competition, the comprehension test, and a questionnaire.

Table 4: Procedures of the experiment

1. Reading full instructions before coming to the experimental lab
2. Reading summarized instructions in the experimental lab
3. Instructive simulation to induce demand and price expectations
4. Comprehension test
5. Trading in the CDA.
    a. Answer questions about their expectations every odd round number.
    b. Trading in the forward market.
        i. Initial phase - 45 seconds
        ii. Conditional phase - min. 30 seconds, max. 180 seconds
    c. Trading in the spot market starts.
        i. Initial phase - 45 seconds
        ii. Conditional phase - min. 30 seconds, max. 180 seconds
        iii. Penalty phase (when retailers that are short of the "units demanded") – max 120 seconds. Each 30 seconds a penalty for each unit that they are short.
        iv. End penalty phase: additional penalty for each unit they are short.
6. Questionnaire
7. Pay-out

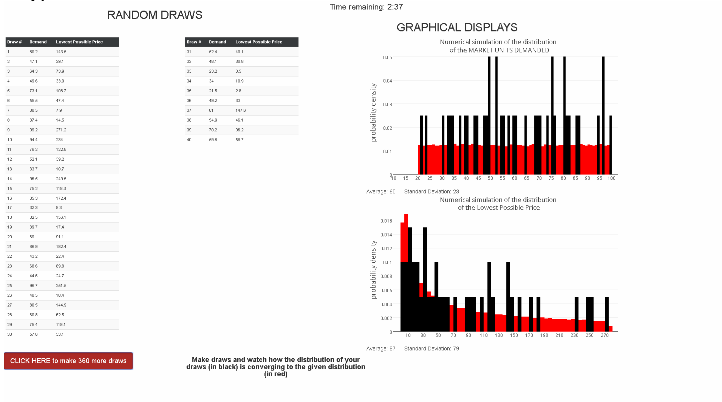
Table 4 gives an overview of the experimental procedures. The procedures were as following:

1. Subjects have to read the full instructions before coming to the experimental lab. See Appendix B3 for an example of the instructions. The full instructions template is coded in the software program.[21] Three days before the experiment, they are send the full instructions. Subjects have to confirm the receipt and subjects that do not confirm are send a reminder every day. Subjects that have not confirmed the receipt of the instructions are not admitted to the experiment.

2. Subject are given the opportunity to refresh their understanding by reading summarized instructions in the experimental lab for maximally 10 minutes. A timer clearly indicates the time left.

[21] The software can be inspected or downloaded, installed and run from https://github.com/slvstr1/dAuction2_public. To run the software, the README note gives extensive instructions, including how to install Vagrant and VirtualBox to create a virtualized Linux Fedora programming and server environment. The statistical files (raw data and STATA do-files) and Appendix B, containing snapshots of the instructive simulation, the questions of the comprehension, and the instructions for treatment 1, can be downloaded from https://github.com/slvstr1/dAuction2_public/tree/master/materials_for_paper.

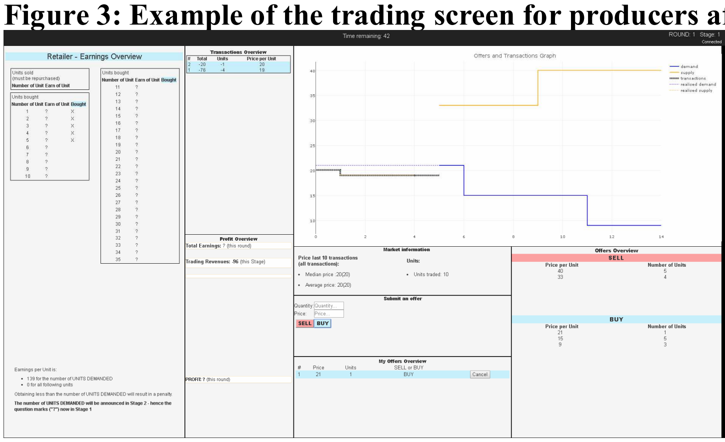
Figure 2: The visual simulation of random draws from a distribution



3. An instructive simulation is presented to subjects to induce demand and price expectations. See Appendix B1 for more snapshots. The simulation shows the drawing of random demand outcomes of a uniform distribution and the effect on market outcomes under perfect competition (See Figure 2). The distribution used in the simulation is the same one as the one used in the treatment.

4. A comprehension test with 19 multiple choice questions and a time limit of 16 minutes is administered to subjects. See Appendix B2 for the questions and answers. A timer clearly indicates the time left. Students are only able to move to the next question when they have answered the question correctly. Out of the 24 students present, the 16 with the fewest mistakes are selected to participate in the experiment. The eight students with the highest number of mistakes receive a show-up fee of CZK 200 (about 8 euro), unless they made more than 16 mistakes, in which case it will be assumed they have not read the instructions and they receive only CZK 50 (about 2 euro).

Figure 3: Example of the trading screen for producers after a specific series of offers.



5. Trading in the CDA commences and lasts for 10 period. Each period proceeds in the following way. Students have to answer questions about their expectation of the market demand, called "units demanded" in the instructions, the average price, and the average price under perfect competition every odd period. Then trading starts in the forward market. Producers and Retailers can make buy or sell offers using a trading screen

(for an example, see Figure 3). In the forward market, retailers nor producers can see what is the number of "units demanded". The forward market initially lasts 45 seconds. A timer clearly indicates the time left. Then a "conditional phase" starts where the market closes if no new transaction is made within 30 seconds. A new transaction sets the timer back to 30 seconds. The conditional phase lasts minimally 30 and maximally 180 seconds.

Once the forward market has ended, the spot market starts. Retailers and producers can now see the number of "units demanded". The spot market initially lasts 45 seconds. Then, as in the forward market, a "conditional phase" starts that lasts minimally 30 and maximally 180 seconds. During the "conditional phase", Retailers and Producers that are short of the units demand are warned about the penalty they may receive if they do not fulfill the "units demanded".

When there are retailers that are short of the "units demanded", the spot market is extended with a "penalty phase" of maximally 120 seconds. Each 30 seconds, the retailers that are units short receive a penalty equal to 10 ECU for each unit that they are short. The moment no retailers are short anymore, the penalty phase ends. If the retailers are still short at the end of the penalty phase, an additional penalty is levied. The additional penalty is equal to the lowest cost that it would take to produce the units by the producers, multiplied by 2.

To avoid the possibility of producers exerting market power due to the (threat of) penalties on retailers, a producer in the spot market during the penalty phase that has sold less than the number of "units demanded" is administered the same penalties as the ones administered to retailers that are short.

6. Subjects fill out a questionnaire.

7. One of the 10 rounds is selected for payout at random and paid out to the subjects.

3.4 Sessions and treatments

Table 5: treatments and independent observations

	Treatment	Independent observations
T1	55-65	12
T2	40-80	12
T3	20-100	12

Twelve data points were collected for each treatment.[22] A total of 432 subjects took part in the experiment, out of which only the 2/3rd best scoring on the comprehension test, 288 subjects, participated in the trading part of the experiment. As each session consists of eight subjects (four as producers and four as retailers), this results in 36 independent observations, evenly distributed over the three conditions.

[22] As discussed earlier, we are agnostic about the precise numerical value of the absolute risk aversion measures for the specific hedging task in this experiment. This makes it impossible to solve Equation 5 for the predicted effect size, thus hampering possible ex-ante power calculations.

4. Results

4.1 Spot prices

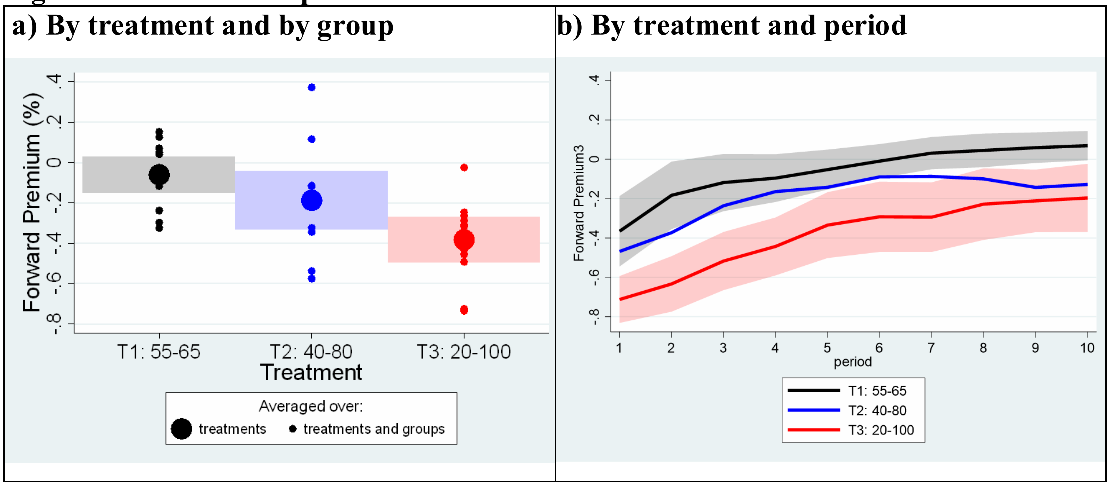
Table 6: Deviation from predicted spot price ($deviation_{spot}$)

Dummy variables	All periods (1-10)	All periods (1-10): Average all treatments	Last periods (6-10)	Last periods (6-10): Average all treatments
constant		0.178* (-0.0958)		0.111* (-0.062)
T1	-0.083** (-0.037)		-0.0072 (-0.035)	
T2	-0.033 (-0.059)		0.038 (-0.066)	
T3	0.65*** (-0.22)		0.30* (-0.16)	
N	360	360	180	180
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1				
F-test on equality (N=360)	**0.0093**		0.1573	
KW on equality (N=36)	0.1503		0.3578	

Table 6 shows the results of analyzing the deviations from the predicted spot prices, calculated as $deviation_{spot} = (p_{spot}^{obs} - p_{spot}^{theory}) / p_{spot}^{theory}$, where p_{spot}^{obs} stands for the observed spot prices averaged over all spot market transactions within one period and p_{spot}^{theory} for the predicted spot price. The deviations are analyzed by running the linear regression $deviation = d_1T_1 + d_2T_2 + d_3T_3 + \varepsilon$ with errors clustered for each group, and by testing for equality of the treatment dummies by running an F-test and also by running a non-parametric test, the Kruskal-Wallis equality-of-populations rank test (KW). The analysis indicates that, on average, the prices in the spot markets deviate from the predicted spot price. In T1 and T2 (the low-variance and medium-variance treatments), the prices are a bit too low, but the deviation is not large (less than 9% and 4% of the theoretical prediction, respectively) and not or only marginally significant. In T3 (the high-variance treatment), the prices are too high by a considerable margin (65%), and the deviation is significant at the 0.05 level. Taken all treatments together, we see that the effect of T3 dominates. The F-test following the regression indicates that the deviations are significantly different between the treatments at the 0.01 level, although the KW indicates insignificance (p=0.15). The prices improve when considering only the last 5 rounds in the experiment. The deviations in T1 and T2 are small (below 1% and 4%) and no longer significant. Also the deviation in T3 decreases, but its size is still considerable (30%) and significant at the 0.1 level. The F-test following the regression and the KW indicate that the deviations are not significantly different between the treatments with p=0.16 and p=0.36, respectively.

4.2 Forward premia

Figure 4: The forward premium



The shaded areas shows the 95% confidence interval for the means (only for T1 and T3 in b).

Figure 4 shows the average values of the relative forward premium. In Figure 4a, the outcomes, for each treatment and independent observation (the group), are averaged over all periods[23] and are shown as small dots. The large dots are the averages over all independent observations within a treatment. The averages show that the relative forward premium becomes more negative with increased demand variance. The larger the standard deviation of demand (20-100), the more negative is the relative forward premium, reaching close to -40% for treatment T3. In contrast, treatment T1, with the small standard deviation of demand (55-65), has a relative forward premium hardly smaller than zero (-6%).

In Figure 4b, the outcomes, for each treatment and period, are averaged over all groups and shown as solid lines. It shows that, in each treatment, the average relative forward premium start somewhat low, and then over periods increases, showing convergence to a higher level. The differences of the averages between the treatments is relatively stable, showing the same ranking (T3 is smaller than T2 is smaller than T1).

The shaded areas around large dots in Figure 4a and around the lines in Figure 4b[24] show the 95% confidence intervals. As indicated by the large shaded areas, the observations show a rather large dispersion, especially for the treatment T2 with the intermediate standard deviation of demand (40-80).

The upper row of Table 7 shows the results of the statistical tests for the relative forward premium. For Hypothesis 1, I use one-sided one-sample Wilcoxon signed rank tests for the null Hypotheses that $FP_i \geq 0$ for $i \in \{T_1, T_2, T_3\}$. Conform the expectations, the null cannot be rejected for the base treatment T1 (the treatment with low demand variance) and can be rejected for treatment T3 (the treatment with the highest demand variance) and the effect is strongly significant ($p<0.001$). The null can also be rejected for treatment T2, the treatment with the intermediate level of demand variance.

[23] Using only the last 5 period leads to qualitatively identical results.

[24] For comprehensibility the confidence intervals are only drawn for treatments T1 and T3.

Table 7. Test results using non-parametric tests

	Hypothesis 1	N	Hypothesis 2	N
Relative Forward Premium	$FP_{T_1} \geq 0$	12	$FP_{T_2} < FP_{T_1}$ *	24
	$FP_{T_2} < 0$ **	12	$FP_{T_3} < FP_{T_2}$ *	24
	$FP_{T_3} < 0$ ***	12	$FP_{T_3} < FP_{T_1}$ **	24

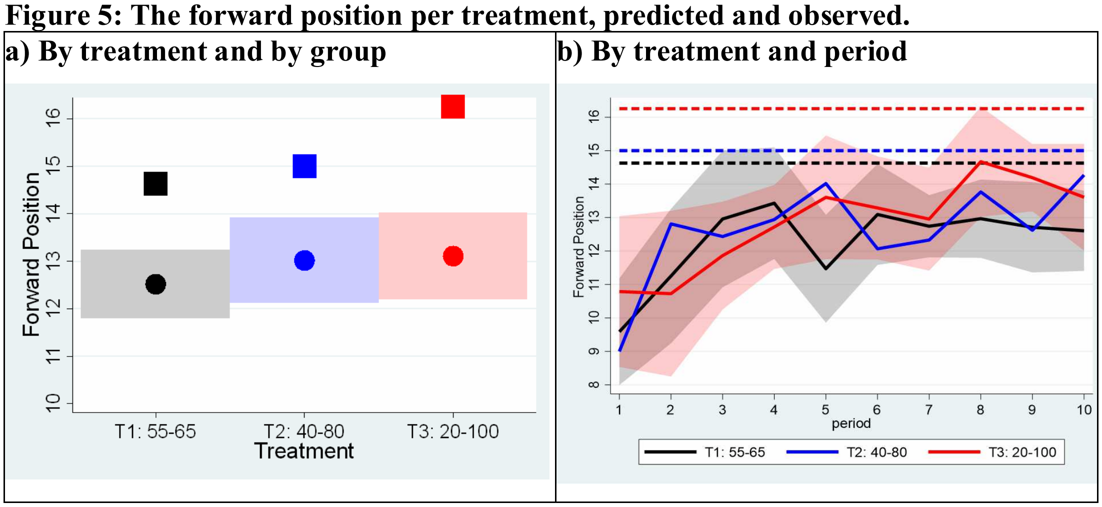
	Hypothesis 3a	N	Hypothesis 3b	N
Forward Positions	$FPos_{T_1} > 0$ ***	12	$FPos_{T_1} \neq 14.6$ ***	24
	$FPos_{T_2} > 0$ ***	12	$FPos_{T_2} \neq 15.0$ ***	24
	$FPos_{T_3} > 0$ ***	12	$FPos_{T_3} \neq 16.3$ ***	24

*** p<0.01, ** p<0.05, * p<0.1

For Hypothesis 2, I use one-sided two-sample Wilcoxon rank-sum (Mann-Whitney) tests for the null Hypotheses that $FP_{T_1} > FP_{T_2}$, $FP_{T_2} > FP_{T_3}$, and $FP_{T_1} > FP_{T_3}$. Conform the expectations, all relationships are significant.[25]

The lower row of Table 7 shows the results of the statistical tests for the forward positions. The forward positions, confirming Hypothesis 3a, are clearly significantly larger than zero. However, disconfirming Hypothesis 3b, they are significantly below the predicted theoretical values.

Figure 5: The forward position per treatment, predicted and observed.



The shaded areas shows the 95% confidence interval for the means (only for T1 and T3 in b).

The conclusions for forward positions are illustrated in Figure 5. In Figure 5a, the outcomes, for each treatment and independent observation (the group), are averaged over all periods and are shown as dots. The squares show the theoretical predictions. The observations are clearly significantly larger than zero and also significantly smaller than the theoretical predictions. Figure 5b illustrates that the forward position improves over time, as the average forward positions come closer to the theoretically predicted values in later periods.

[25] As a robustness test, a Jonckheere–Terpstra test was used. The test rejected any possible alternative order of treatments with a significance of p=0.03. Further robustness tests can be found in Appendix A4, using linear regressions with clustered errors.

4.3 Observed (implied) levels of absolute risk aversion

Table 8. Implied level of absolute risk aversion

	All periods (1-10)		Last periods (6-10)	
Dummy variables		Average all treatments		Average all treatments
Constant		0.037 (0.022)		-0.017 (0.020)
T1	0.095 (0.064)		-0.060 (0.060)	
T2	0.012** (0.0044)		0.0067 (0.0043)	
T3	0.0030*** (0.0004)		0.0019*** (0.0006)	
N	360	360	180	180
	Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1			
F(2,35)-test on equality (N=360)	**0.06**		0.32	
KW on equality (N=36)	0.13		**0.04**	

The estimates for the forward premia can be used to recover the observed or implied level of absolute risk aversion. Manipulating Equation (5), the observed absolute risk aversion is given by $RA^{obs} = RA^{theory} \cdot (FP^{obs} / FP^{theory})$.[26] Table 8 shows the results of analyzing the implied level of absolute risk aversion. They are analyzed by running the linear regression $RA^{obs} = d_1T_1 + d_2T_2 + d_3T_3 + \varepsilon$ with errors clustered for each group, and by testing for equality of the treatment dummies by running, following the linear regression, an F-test and a non-parametric test, the Kruskal-Wallis equality-of-populations rank test (KW). The F-test and KW are barely in agreement. For all periods, the F-test indicates that the implied levels of absolute risk aversion are significantly different over the treatments, but weakly so (p=0.06), while the KW indicates indifference at a level of (p=0.13). In the last 5 periods, the F-test indicates that the implied levels of abosulute risk aversion are not significantly different (p=0.32), while the KW indicates a significant difference (p=0.04). Given the few independent observations and considerable heteroscedasticity, I tend to judge the KW test as more reliable, thus indicating differences in the level of absolute risk aversion between the treatments.

Indeed, the averages suggest that the coefficient of absolute risk aversion may vary from 0.002 for treatment T3 to 0.06 for treatment T1. The effect is that the observed differences between the forward premium are less than predicted, but not strong enough to neutralize the predicted effects, as shown in Figure 4and the statistical tests in Table 7.

---

[26] Using equation 5, amd writing $k^{theory} = -\dfrac{N_P}{(N_P + N_R)}\dfrac{1}{ca^x}\left(c\overline{P}_R \,\mathrm{Cov}[P_S^x, P_S] - \mathrm{Cov}[P_S^{x+1}, P_S]\right)$, then for both observed and theoretical forward premium, it must hold that $FP^{theory} = RA^{theory} \cdot k^{theory}$ and $FP^{obs} = RA \cdot k^{theory}$. Then, by dividing the two equations and rearranging results in $RA^{obs} = RA^{theory} \cdot \dfrac{FP^{obs}}{FP^{theory}}$.

5. Conclusion

In this paper, I perform a first experimental test of Bessembinder and Lemmon (2002) (BL). I create treatments with different demand distributions to establish, in an experimental setting, a market premia effect. Indeed, by increasing the variance of demand, while holding mean demand constant, the forward premium becomes, as predicted by the theory, more negative. In the experiments, the spot prices are by and large close to the predicted levels in the treatments with no or low variance (T1 and T2), but are a bit higher in the treatments with high variance (T3). The experimental support is a qualified but important contribution, as the empirical evidence on the validity of BL's theory is mixed.

The experimental data also suggest that the level of absolute risk aversion does not stay constant, but rather decreases when the variance of demand increases. This seems intuitive, as with a higher variance of demand, prices and the retail price increase. While the change in absolute risk aversion somewhat lowers the differences in forward premia between the treatments, it is by far not enough to neutralize these effects.

In the experiment, special care has been taken to keep the trading setup tractable for the experimental participants, by using a dynamic graphical presentation of the order book and by using simple (uniform) market demand distribution. Also, special care has been taken to assure the participants' correct understanding of the intricate problem embedded in the experiment by presenting not only instructions, but also an instructive simulation and test questions. In addition, only the 2/3th of the participants with the best scores of the test questions within a session were selected to take part in the experiment.

The experimental support is qualified as an experimental test is – necessarily by design – operationalized for specific parameterizations and implementation details. It is thus worth repeating that a specific type of distribution is used, with specific parameter values for the distribution, and the cost functions. Also, the trading has been implemented in a specific trading environment, the continuous double auction market (CDA).

Future studies can build upon the results of this first experimental study by testing BL's theory for different parameterizations and implementation details. Naturally, different types of distributions, different parameter values for the distribution or cost functions and different trading environments than the CDA (for example, the clearinghouse auction) come to mind. In addition, parameter values could be chosen, such that the theory predicts also positive forward premia. Studying positive forward premia will, however, require special care as numerical simulations numerical indicate that this makes it very likely for retailers to make negative profits.

6. References

Bessembinder, H., Lemmon, M.L. (1992) Systematic risk, hedging pressure, and risk premiums in futures markets. Journal of Finance 57 (3), 1347–1382.

Botterud, A., Kristiansen, T., & Ilic, M. 2010. The relationship between spot and futures prices in the Nord Pool electricity market. Energy Economics 32 (5), 967-978.

Bunn, D.W., Chen, D. 2013. The forward premium in electricity futures. Journal of Empirical Finance 23, 173–186.

Cason, T. N., & Friedman, D. (1996). Price formation in double auction markets. Journal of Economic Dynamics and Control, 20(8), 1307-1337.

Cason, T. N., & Friedman, D. (2008). A comparison of market institutions. Handbook of experimental economics results, 1, 264-272.

Daskalakis, G., & Markellos, R. N., (2009) Are electricity risk premia affected by emission allowance prices? Evidence from the EEX, Nord Pool and Powernext. Energy Policy 37, 2594–2604.

Diko, P., Lawford, S., Limpens, V., 2006. Risk premia in electricity forward prices. Studies in Nonlinear Dynamics & Econometrics 10(3) (Article 7).

Douglas, S., Popova, J., 2008. Storage and the electricity forward premium. Energy Economics 30, 1712–1727.

EU Commission, 2014. Communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions. A policy framework for climate and energy in the period from 2020 to 2030. Brussels, 22.1.2014 COM(2014) 15 final.

EEX 2005, Annual Report 2005. Available at https://www.eex.com/en/about/eex/annual-report.

EEX 2019, Annual Report 201. Available at https://www.eex.com/en/about/eex/annual-report.

Fleten, S. E., Hagen, L. A., Nygård, M. T., Smith-Sivertsen, R., & Sollie, J. M. (2015). The overnight risk premium in electricity forward contracts. Energy Economics, 49, 293-300.

Furio, D., & Meneu, V. (2010). Expectations and forward risk premium in the Spanish deregulated power market. Energy Policy, 38, 784–793.
Fleten, S. E., Hagen, L. A., Nygård, M. T., Smith-Sivertsen, R., & Sollie, J. M. (2015). The overnight risk premium in electricity forward contracts. Energy Economics, 49, 293-300.
Hadsell, L., Shawky, H.A., 2006. Electricity price volatility and the marginal cost of congestion: an empirical study of peak hours on the NYISO market, 2001–2004. The Energy Journal 27(2), 157–179.

Handika, R., Trueck, S. 2013. Risk Premiums in Interconnected Australian Electricity Futures Markets. SSRN Working Paper (http://dx.doi.org/10.2139/ssrn.2279945).

Haugom, E., Ullrich, C.J., 2012. Market efficiency and risk premia in short-term forward prices, Energy Economics 34(6), 1931-1941.

Hirshleifer, D. 1990 Hedging Pressure and Futures Price Movements in a General Equilibrium Model. Econometrica 58(2), 411-428.

Hoenig, J. M., & Heisey, D. M. (2001). The abuse of power: the pervasive fallacy of power calculations for data analysis. The American Statistician, 55(1), 19-24.

Jacobs, K., Li, Y., & Pirrong, C. (2017). Supply, Demand, and Risk Premiums in Electricity Markets. Available at SSRN 3066456.

Jones, K. (2018). An examination of prices on the miso exchange. In Global Tensions in Financial Markets (pp. 57-73). Emerald Publishing Limited.

Karakatsani, N.V., Bunn, D.W., 2005. Diurnal Reversals of Electricity Forward Premia. Working paper. London Business School.

Laporte, S., Divine, M., Girault, D., Boutouyrie, P., Chassany, O., Cucherat, M., de Trogoff, H., Dubois, S., Fouret, C., Hoog-Labouret, N. and Jolliet, P. (2016). Clinical research and methodology: What usage and what hierarchical order for secondary endpoints?. Therapie, 71(1), 35-41.

London_Economics, 2012. Energy Retail Markets Comparability Study. Available on: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/200622/London_Economics_-_Energy_Retail_Markets_Comparability_Study.pdf

Longstaff, F.A.,Wang, A.W., 2004. Electricity forward prices: a high-frequency empirical analysis. Journal of Finance 59 (4), 1877–1900.

Lucia, J. J., & Torro, H. (2008). Short-term electricity futures prices: Evidence on the time-varying risk premium. Working Paper Series EC 2008-08.

Karakatsani, V.N. Bunn, D.W. 2008. Intra-day and regime-switching dynamics in electricity price formation. Energy Economics 30, 1776-1797.

McCabe, K., Rassenti, S., Smith, V. (1993). “Designing a uniform price double auction: An experimental evaluation”. In: Friedman, D., Rust, J. (Eds.), The Double Auction Market. Addison–Wesley, Reading, MA, pp. 307–332.

Newbery, D. M. (1988). On the accuracy of the mean-variance approximation for futures markets. Economics letters, 28(1), 63-68.

Friedman, D., & Ostroy, J. (1995). Competitivity in auction markets: An experimental and theoretical investigation. The Economic Journal, 105(428), 22-53.

Kagel, J. and D. Levin. (2015). “Auctions: A survey of experimental research.” In The Handbook of Experimental Economics, edited by Kagel, J. and A. Roth. (Eds.) Vol. 2, 563–637. New Jersey, USA: Princeton University Press.

Monitoring Analytics, 2015. PJM State of the Market – 2015. Available at: monitoringanalytics.com/reports/PJM_State_of_the_Market/2015/2015q2-som-pjm-sec3.pdf

Redl, C., Bunn, 2013. Determinants of the premium in forward contracts. Journal of Regulatory Economics 43, 90–111.

Redl, C., Haas, R., Huber, C., & Böhm, B. (2009). Price formation in electricity forward markets and the relevance of systematic forecast errors. Energy Economics, 31(3), 356–364.

Smith, V. L., Williams, A. W., Bratton, W. K., & Vannoni, M. G. (1982). Competitive market institutions: Double auctions vs. sealed bid-offer auctions. The American Economic Review, 72(1), 58-77.

Staffell, I., & Pfenninger, S. (2018). The increasing impact of weather on electricity supply and demand. Energy, 145, 65-78.

Van Koten, S. and Ortmann, A. 2008. The unbundling regime for electricity utilities in the EU: A case of legislative and regulatory capture? Energy Economics 30(6), 3128-3140.

Van Koten, S. and Ortmann, A. 2013. Structural versus Behavioral Remedies in the Deregulation of Electricity Markets: An Experimental Investigation Guided by Theory and Policy Concerns. European Economic Review 64, 256-265.

Van Koten,S. 2020. Forward Premia in Electricity Markets: A replication study. Manuscript submitted for publication.

Weron, R. (2008). Market price of risk implied by Asian-style electricity options and futures. Energy Economics, 30(3), 1098–1115

Weron, R., & Zator, M. (2014). Revisiting the relationship between spot and futures prices in the Nord Pool electricity market. *Energy Economics*, *44*, 178-190.

Xiao, Y., Colwell, D. B., & Bhar, R. (2015). Risk premium in electricity prices: evidence from the PJM market. Journal of Futures Markets, 35(8), 776-793.

Zator, M., 2013. Relationship between spot and futures prices in electricity markets: Pitfalls of regression analysis. HSC Research Reports HSC/13/06, Hugo Steinhaus Center, Wroclaw University of Technology.

7. Appendix A

A1. Derivation of hypotheses 1 and 2

$$FP = -\frac{N_P}{\left(N_P + N_R\right)} \frac{RA}{ca^x}\left(c\overline{P}_R \operatorname{Cov}[P_S^x, P_S] - \operatorname{Cov}[P_S^{x+1}, P_S]\right). \tag{5}$$

BL approximate Equation (5) using a Taylor approximation to obtain a formula of the forward premium as a function of the anticipated variance and skewness of spot prices:

$$\begin{aligned} &FP = b_1 \operatorname{Var}[P_S] + b_2 \operatorname{Skew}[P_S] \text{, where:} \\ &b_1 = \frac{N_P(x+1)\cdot RA}{\left(N_P + N_R\right)ca^x}\left[\mathrm{E}[P_S]^x - P_R\,\mathrm{E}[P_S]^{x-1}\right] \text{ and } b_1 < 0 \\ &b_2 = \frac{N_P(x+1)\cdot RA}{2\left(N_P + N_R\right)ca^x}\left[x\,\mathrm{E}[P_S]^{x-1} - (x-1)P_R\,\mathrm{E}[P_S]^{x-2}\right] \text{ and } b_2 > 0 \end{aligned} \tag{A1}$$

BL thus predict that the forward premium will decrease in the anticipated price variance (as $b_1 < 0$) and increase in the anticipated price skewness (as $b_2 > 0$). An intuitive explanation is that an increase in the wholesale price variance will increase the profit variance of both producers and retailers, but less so for retailers. The profit variance increases less for retailers as the profit-increasing effect of higher demand tempers the profit-decreasing effect of a higher wholesale price. As producers now bear the highest risk, they are eager to hedge by selling units forwards, resulting in a lower forward price and thus a negative forward premium.

An increase in the wholesale price skewness, resulting in more frequent price spikes, will increase the variance of profits of both producers and retailers, but more so for retailers. With a price spike, the wholesale price will be higher that the retail price, leading to losses. Therefore, for the retailer, both the effect of a higher wholesale price and the effect of a higher demand are now strongly profit-decreasing. As retailers now bear the highest risk, they are eager to hedge by buying units forwards, resulting in a higher forward price and thus a positive forward premium.

A2. Simulations based on the theory[27]

As in Bessembinder & Lemmon (2002), I use Equation 5 to run simulations. using the parameter values in Table A1, I calculate outcomes (spot prices, forward prices, forward premia and optimal forward positions) for different demand distributions.

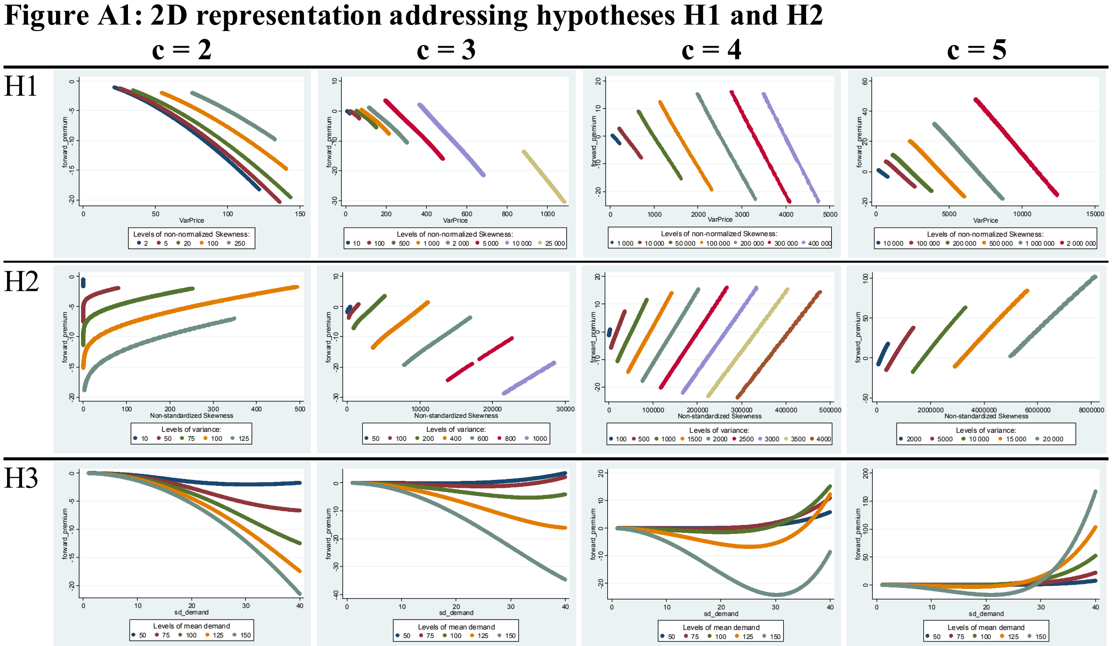
Table A1: Parameters of the data generation process

Retail rate setting methods	$P_R = 1.2 \cdot \frac{1}{N}\sum_i^N P_{S_i}$ *
Risk aversion *RA*	$0.8 \cdot 2^{-c}$ *
Numbers of retailers *NR*	20*
Numbers of producers *NP*	20*
Number of configurations	**4**
Cost convexity parameter *c*	2, 3, 4*, 5
Number of distributions per configuration	**195,891**
Range of demand standard deviation	1-40* (391 steps of 0.1)
Range of mean demand	50 – 150 (501 steps of 0.2)
Scaling	No scaling* ($a = 30(N_P/100)^{c-1}$)
Sample size per distribution	10 001

* Identical to the parameters applied in the simulations used to motivate hypotheses H3 and H4 in Bessembinder & Lemmon (2002).

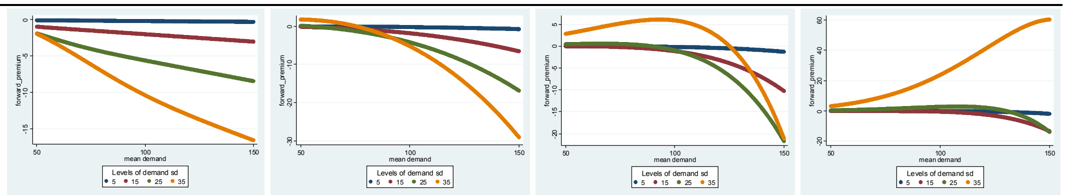
I show here the outcomes for four configurations, one each for a different value of the cost convexity parameter (2, 3, 4, and 5). For each configuration, outcomes are calculated for 195,891 different demand distributions (391 values for the demand standard deviation and 501 values for the mean demand). For each demand distribution, I use a grid spanning 10 standard deviations with 1000 points per standard deviation. Demand realizations that are negative are disregarded, as negative demand is not accounted for in the theory. I use the data to create two-dimensional plots.

Figure A1: 2D representation addressing hypotheses H1 and H2



[27] The simulations and results are based on Van Koten (2020).

H4



The data shown in the form of plots in Figure A1 support hypotheses H1, H2, and H3, but not H4. In line with Hypothesis H1, the forward premium is decreasing in the anticipated price variance (row H1). In line with Hypothesis H2, the forward premium is increasing in the anticipated price skewness (row H2). In line with Hypothesis H3, the forward premium is first decreasing and then increasing in demand variance. In contradiction to Hypothesis H4, an increase in mean demand can result in a lower forward premium for some ranges of parameter values (row H4). See Van Koten (2020) for a further discussion.

A3. Robustness Tests

Table A2. Linear regression results

fpremium_relative	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]
T1	-0.062	0.042	-1.49	0.145	-0.147	0.022
T2	-0.193	0.072	-2.70	0.011	-0.338	-0.048
T3	-0.386	0.055	-6.97	0.000	-0.499	-0.274
Mean dependent var		-0.214	SD dependent var			0.326
R-squared		0.419	Number of obs			360
F-test		19.377	Prob > F			0.000
(Std. Err. adjusted for 36 clusters)						

As a robustness test, I also test the effect of demand variance on forward premia using linear regressions. Table A2 shows the results of a linear regression of the relative forward premium on the treatment, $FP = d_1T_1 + d_2T_2 + d_3T_3 + \varepsilon$, adjusting the standard errors for clustering on the group. The linear regression thus exploits more of the available information than the non-parametric tests. It uses the average for each period for each group., whereas the non-parametric tests average over all periods for each group and thus uses 120 observations per treatment, whereas the nonparametric test uses 12 observations per treatment. The results are virtually identical.

Table A3. Test results using parametric tests.

Hypothesis 1	Significance	Hypothesis 2	Significance
$FP_{T_1} \geq 0$	*p=0.078*	$FP_{T_2} < FP_{T_1}$	*p=0.061*
$FP_{T_2} < 0$	*p=0.006*	$FP_{T_3} < FP_{T_2}$	*p=0.020*
$FP_{T_3} < 0$	*p<0.001*	$FP_{T_3} < FP_{T_1}$	*p<0.001*

N= 360, Independent clusters = 36

Using the linear regression, the difference of the dummies from zero are used to test Hypothesis 1 and additional F-tests are run to test Hypothesis 2. Table A3 shows the results. They are virtually identical to the ones using non-parametric tests as shown in Table 6.

The result are in line with those in section 4 and thus support the theoretical predictions of BL: An increase in the variance of demand indeed lowers the forward premium for the specific parameter values used in this experiment.

A4. Competitiveness of Electricity Markets

Electricity markets around the world are generally not perfectly competitive. Regarding producers, a study by London Economics (2012) shows that out of 57 regions, including EU countries, US states, New Zealand, the HHI ranges from 510 to 6445 with 2075 the median. The same range of HHI values is obtained when symmetrical producers range between 2 and 25 with 5 the median. For the EU15 the HHI ranges in 2010 from

400 to 10000, with 3000 the median. The same range of HHI values is obtained when symmetrical producers range between 1 and 25 with 3 the median. The NMS12 are known to be less competitive and have higher HHI values. While the numbers used for the EU are not very recent, the market share of the largest generator in the electricity market has changed very little from 2004 till 2013 according to Eurostats data[28]. The PJM market shows in the first half of 2015 depending on the specific hour, that the HHI ranges between 916 and 1468 (Monitoring Analytics, 2015). The same range of HHI values is obtained when symmetrical producers range between 7 and 10.

The competitiveness of retail markets is very hard to determine due to data problems. A study by London Economics (2012) estimates that in the same 57 regions, the number of retailers (also referred to as "suppliers") with a market share above 5% ranges from 1 to 10, with the average for EU being 4 and for US 5. However, these figures may underestimate the number of retailers, as retailers with a market share below 5% are excluded. The number may also, however, overestimate the competitiveness in retail on the national level, as some retailers sell only in certain areas and may thus have a larger amount of market power than suggested by the number of retailers in the national market.

A5. Parameters

Production Cost

MC[1]= 0.02	MC[7]= 6.	MC[13]= 39	MC[19]= 120	MC[25]= 280	MC[31]= 530
MC[2]= 0.14	MC[8]= 9.	MC[14]= 49	MC[20]= 140	MC[26]= 310	MC[32]= 580
MC[3]= 0.48	MC[9]= 13	MC[15]= 60	MC[21]= 165	MC[27]= 350	MC[33]= 640
MC[4]= 1.1	MC[10]= 18	MC[16]= 75	MC[22]= 190	MC[28]= 390	MC[34]= 700
MC[5]= 2.2	MC[11]= 24	MC[17]= 85	MC[23]= 215	MC[29]= 430	MC[35]= 760
MC[6]= 3.8	MC[12]= 31	MC[18]= 105	MC[24]= 245	MC[30]= 480	

In the calculations, the cost function $C_i = F + \frac{a}{c}\left(Q_{Pi}\right)^c$, with $a = \frac{4}{225}$, $F = 0$, and $c = 4$ is used, and the marginal costs are calculated as $MC[0] = 0$ and $MC[q] = C[q] - C[q-1]$.

[28] nrg_ind_331a at http://ec.europa.eu/eurostat/data/database, accessed on 2015.09.24.