2-1-2017

TAX INCIDENCE: DO INSTITUTIONS MATTER? AN EXPERIMENTAL STUDY

James Cox
Georgia State University

Mark Rider
Georgia State University

Astha Sen
Sonoma State University

Tax Incidence: Do institutions matter? An experimental study

By James C. Cox[a], Mark Rider[a], and Astha Sen[b]

Forthcoming in *Public Finance Review*

[a] Experimental Economics Center and Department of Economics, Andrew Young School of Policy Studies, Georgia State University.

[b] School of Business and Economics, Sonoma State University.

Corresponding author: James C. Cox, office phone (404) 413-0200 and email address jccox@gsu.edu.

Abstract

According to economic theory, the incidence of a unit tax is independent of the statutory assignment of the liability to pay the tax. However, the theory is silent on the possible effects of market institutions on tax incidence. We report data from an experiment designed to address two questions. Is tax incidence independent of the assignment of the liability to pay tax to sellers or to buyers? Is tax incidence independent of market institutions? We conduct laboratory experiments with double auction and posted offer markets. Based on the results of nonparametric and parametric tests of prices generated by laboratory markets, we conclude that the answer to both questions is "no". We report that observed differences from liability side equivalence are statistically significant and economically meaningful. We also report that the incidence of the same tax differs between double auction and posted offer markets with the same demand and supply schedules.

Keywords: tax incidence, market institutions, experiment

JEL codes: C9, D4, H2

Introduction

To understand the distributional effects of a tax, it is necessary to know who ultimately bears its burden. The theory of tax incidence is concerned with answering this very question, and there may not be a more important one in the field of public finance. According to the liability side equivalence (LSE) hypothesis, the economic incidence of a unit tax is independent of the statutory assignment of liability to pay the tax when markets are perfectly competitive. Whether the tax is on the sellers or on the buyers in the market, the economic incidence of a tax is hypothesized to depend on the relative elasticities of supply and demand: the more inelastic of the two bears the greater burden.[1] Moreover, Weyl and Fabinger (2013) show that symmetric markets satisfy LSE even when they are imperfectly competitive.

Since governments rarely change the statutory assignment of the liability to pay tax, there are few opportunities to test the LSE hypothesis using observational data.[2] However, there is a growing literature showing that behavioral, informational, and institutional factors influence the incidence of a tax.[3] Markets need institutions to function. These institutions specify how buyers and sellers interact to determine prices and quantities. Different market institutions are known to have different price formation and quantity determination properties, and these properties may affect the incidence and excess burden of taxes.[4]

Nevertheless, tax incidence theory does not specify market institution(s), so it is unclear what form of market exchange may or may not exhibit LSE. Our central questions are whether the predictions of the theory are consistent with data from exchange in laboratory versions of two prominent market institutions, double auction (DA) and posted offer (PO) markets. To gain insight into the effect of market institutions on tax incidence, we conduct experiments comparing computerized DA and PO markets.

The rules of the DA market in our experiments are a simplified version of those that govern trading on the New York Stock Exchange and on many other stock, commodities, and futures markets. An experimental DA market is open for a specified interval of time; buyers are free to announce bid prices for the abstract commodity they wish to buy and sellers are free to announce ask prices for the commodity they wish to sell. In the simplified DA markets used in many experiments, including ours, each bid, ask, or contract is for a single unit. Actual trade occurs when a seller accepts the most attractive bid price, or a buyer accepts the most attractive ask price, among the outstanding bids and asks.

The rules of the PO market in our experiments resemble those that govern most of the consumer goods markets in developed countries. Think, for example, of a supermarket or department store. Sellers post prices for a commodity and may also limit the quantities they are willing to sell to "stock on hand". Buyers make the decision to buy or not to buy from alternative sellers at their posted prices.[5] The computerized PO markets in our experiments are simplified versions of this field institution.

In addition to testing LSE, we also pose and attempt to answer a new question: Is tax incidence independent of the market institution? In other words, for a given assignment of the liability to pay tax, does tax incidence vary between market institutions? Given the current state of tax-incidence theory, we continue to assume that tax incidence depends solely on the relative elasticities of supply and demand. Applying this assumption in the current context implies that tax incidence should be independent of market institutions, especially when markets are symmetric. For ease of reference, we henceforth refer to this prediction as the independence of market institutions (IMI) hypothesis. This question is not addressed in the previous literature on tax incidence. This illustrates the fact that the public finance literature has not done an adequate job of grappling with the extensive literature on market institutions. Given the practical importance of tax incidence to tax policy design, it is difficult to overstate the importance of investigating the implications of market institutions for tax incidence.

Our baseline experimental design includes four treatments that use the same symmetric supply and demand schedules: (1) a double-auction market with a unit tax on the buyer (DATB); (2) a double-auction market with a unit tax on the seller (DATS); (3) a posted-offer market with a unit tax on the buyer (POTB); and (4) a posted-offer market with a unit tax on the seller (POTS). We report several nonparametric and parametric tests of the LSE and IMI hypotheses. We also estimate Random Effects (RE) models with bootstrap clustered standard errors to test these hypotheses. In the four baseline treatments described above, the tests reject both the LSE and IMI hypotheses at conventional levels of statistical significance. Furthermore, we report evidence that deviations from LSE are economically meaningful for all but the DATB treatment.

We are not the first to propose testing the LSE hypothesis in a laboratory setting. Kachelmeier, Limberg, and Schedewald (1994), Kerschbamer and Kirchsteiger (2000), Borck et al. (2002), Riedl and Tyran (2005), and Ruffle (2005) test the LSE hypothesis in laboratory settings using a variety of tax types and institutions. We provide a more elaborate review of the literature in the Online Supplement. The study by Borck et al. (2002) most resembles our approach. They test the LSE hypothesis in DA and PO markets, using a unit tax. They fail to reject the LSE hypothesis for either market institution. They do not report whether tax incidence differs between the two market institutions, which is one of our central questions. We compare our findings with theirs in greater detail in the conclusion to this paper.

The remainder of this paper is organized as follows. The next section provides a conceptual framework, and section 3 describes our experimental design and protocol. Section 4 explains the baseline results. In section 5, we examine the robustness of our conclusions to alternative experimental designs and discuss the external validity of our main results. Section 6 concludes.

Price formation in market institutions

DA and PO market institutions are known to have different price formation and quantity determination properties. These differing properties may affect the incidence and excess burden of taxes. Smith (1976b), Williams (1980), Smith et al. (1982), and Smith and Williams (1983) report the robust result that DA markets converge rapidly to competitive equilibrium quantities, thus exhausting the potential gains from trade. Although DA markets do not operate in the manner described by the model of perfect competition, they do achieve the Pareto efficient allocation of competitive market theory. Also, price convergence in DA markets with symmetric demand and supply schedules is unbiased, neither converging consistently from above or below (Smith and Williams 1982). Therefore, the DA market institution with symmetric demand and supply schedules seems like a promising candidate to exhibit LSE.

Plott and Smith (1978), Williams (1980), Hong and Plott (1982), Smith (1982b), and Ketcham, Smith, and Williams (1984) report that PO markets produce prices that converge to an equilibrium, although not necessarily a competitive one, and also yield less efficient allocations than DA markets. Consequently, PO markets may not exhibit LSE. PO exchange, however, is the most common retail market institution in advanced market economies; thus, it is important to investigate the effect of this market institution on tax incidence.

Finally, given the different efficiency and price-convergence properties of the DA and PO market institutions, it would not be surprising if the data reject the IMI hypothesis.

Experimental design and protocol

We proceed in this section by describing our experimental design and protocol; then we discuss subject responses to the questionnaire and subject earnings in the four baseline treatments.

Experimental design

Following Smith (1976a and 1982a), we induce stationary demand and supply functions for a fictitious commodity. There are five sellers and five buyers in any given market. Each of the five sellers is assigned a cost of twenty-three experimental dollars (ED) for unit 1, twenty-six ED for unit 2, twenty-nine ED for unit 3, thirty-two ED for unit 4, thirty-five ED for unit 7, and thirty-eight ED for unit 6. These marginal costs identify the induced individual supply schedules used in the experiment. Each of the five buyers is assigned a value of fifty ED for unit 1, forty-seven ED for unit 2, forty-four ED for unit 3, forty-one ED for unit 4, thirty-eight ED for unit 5, and thirty-five ED for unit 6. Figure A.1 in the Online Supplement illustrates the induced demand and supply schedules used throughout every treatment reported in this study.

The symmetric market design is the discrete-variables analogue of a continuous-variables model of a market in which the own-price elasticity of demand equals the own-price elasticity of supply. The induced demand and supply schedules are symmetric in several ways. First, the market demand and supply schedules create equal buyer and seller surpluses at the competitive equilibrium quantity. This is important because unequal buyer and seller surpluses can produce biased convergence paths even in a DA market (Smith and Williams 1982) and that may affect tax incidence. Secondly, the symmetry of the market demand and supply schedules implies equal demand and supply price elasticities. This supports clear tests of empirical incidence compared to a theoretical prediction of equal sharing of the tax burden. A third way in which the design is symmetric is across individual buyers and sellers. Each buyer (respectively seller) has a unit at each of the induced demand (respectively supply) prices. This provides equal payoff incentives across all subjects in the experiment.

In the absence of a tax, the competitive equilibrium price tunnel is between thirty-five and thirty-eight ED, and the unique competitive equilibrium quantity is twenty-five units, consisting of five units traded by each of the five buyers and sellers. With a unit tax of twelve ED, the competitive equilibrium price tunnel is between twenty-nine and forty-four ED, and the unique competitive equilibrium quantity is fifteen units, consisting of three units traded by each of the five buyers and sellers. The quantity reduction from twenty-five to fifteen units implies an excess burden of the tax equal to sixty ED [$= 5\times(38 - 35) + 5\times(41 - 32)$].

To gauge the impact of market institutions on tax incidence, we change the type of market institution from a DA to a PO market, keeping the assignment of the liability to pay tax the same. Similarly, we change the assignment of the liability to pay tax from the seller to the buyer, keeping the market institution the same, to gauge the impact of this change on the incidence of a tax. The result is a two-by-two design with four treatment cells.

We conduct a total of four sessions, where each session is devoted to one of our four treatment cells. At the beginning of each session, the subject's role as a buyer or a seller is randomly assigned by a computer and remains the same throughout the session. In each session, four independent markets of the same type, consisting of five buyers and five sellers in each market, are simultaneously trading. Each buyer and seller is given five (no-tax) infra-marginal units to buy or sell at the beginning of each trading period, and there are thirty trading periods in each of the four markets in a session. In short, the experiment uses a between-subjects design, with forty subjects randomly assigned to the role of buyer or seller in equal numbers in each of the four treatments, resulting in a total of 160 subjects.

Experimental protocol

At the beginning of each experimental session, subjects read detailed instructions appearing on their computer screens on how to interact with the computer to trade in the market. After the subjects read the instructions, summary instructions are projected on a screen and read to the subjects to increase understanding of the market participation process. (Subject instructions are available at http://expecon.gsu.edu/jccox/subjects.html.) Before actual trading periods begin, there are five practice trading periods. These practice periods acquaint the subjects with the software and help them to understand the decision-making process. Subjects are permitted to ask questions of the experimenter during written instruction and oral summary parts of a session and during practice trading periods. After the practice periods, there are thirty actual trading periods in each session. The total number of trading periods is not announced to the subjects.[6]

Induced value and tax information given to subjects

In the four baseline treatments, subjects are given the same information about induced marginal costs and values. Costs and values are private information. Throughout each session, the subjects are seated in a manner that protects the privacy of this information. The amount of the unit tax (twelve ED) and the assignment of the liability to pay tax are clearly stated in both the detailed and summary subject instructions. The costs, values, and tax per unit remain the same throughout each session.

The tax as well as costs and values are made salient to the subjects. A seller's trading screen shows "cost per unit" in one column and the "cost plus tax per unit" in an adjacent column.[7] When the buyer is assigned the liability to pay the tax, the figures in these two columns are identical because the unit tax on the seller is equal to zero. Similarly, a buyer's trading screen shows "value per unit" in one column and the "value minus tax per unit" in an adjacent column. When the seller is assigned the liability to pay the tax, the figures in these two columns are identical, because the unit tax on the buyer is equal to zero. Further details on the exact layout of the buyer and seller trading screens are shown in the screen shots of the subject instructions in the URL above. .

Double auction markets

Two baseline treatments are conducted with DA markets. In one baseline treatment, the liability to pay tax is assigned to sellers and, in the other treatment, to buyers. The assignment of the liability to pay tax is the only difference in the two baseline DA treatments. The unit tax in both treatments is twelve ED. In both treatments, sellers and buyers are given two and a half minutes to complete their transactions in each trading period. The time remaining in the trading period, the subject's own number of units available to buy or sell, own cumulative earnings, own profit or loss from each traded unit, outstanding bid and ask prices in the market, tax charged per unit, and the market transaction prices are displayed on the buyers' and sellers' trading screens throughout a trading period. The bid, ask, and transaction price information provided on a subject's computer screen is a characteristic of the DA market institution in the field.

Posted offer markets

The remaining two baseline treatments are PO markets. In this market institution, sellers make the first move by posting an offer price and the number of units they are willing to sell at that offer price. Buyers then enter the market in a random queue, one by one, and accept the sellers' offers if they find them attractive. If an offer is accepted, then trade occurs. In one baseline PO treatment, the liability to pay the unit tax is assigned to sellers; in the other baseline treatment, it is assigned to buyers. The assignment of the liability to pay tax is the only difference between these two baseline treatments. Sellers have two and a half minutes to post offers in each trading period. Each buyer has thirty seconds to accept (or not) one or more available offers. The time remaining to make decisions, own number of units available to buy or sell, own cumulative earnings, own profit or loss from each traded unit, and the unit tax are displayed on the sellers' and buyers' trading screens throughout a trading period.

In the baseline treatments of the PO market, a seller's trading screen only lists the offers posted by oneself. Sellers are not able to see the offers posted by other sellers in the market. We believe that this feature of the treatment best reflects the field institution that we are trying to replicate in the laboratory. In retail markets, for example, the PO market institution *itself* does not provide information on the prices posted by other sellers in the way that (bid and) ask prices are publicly provided to all sellers (and buyers) in DA markets. While it is true that retail stores can conduct price surveillance and market studies, conducting such studies makes the point that the market institution itself does not provide pricing information to other sellers in the market. As one of our robustness checks, which are discussed in Section 5, we also report the results of PO treatments in which sellers observe the posted price-quantity pairs of other sellers. This change in the experimental design has no effect on our conclusions.

Because our choice of the baseline information conditions for PO markets differs from many PO market experiments reported in the existing literature, we provide a discussion of this experimental design feature in the Online Supplement.

A buyer's trading screen lists the number of units available for sale and the offer price corresponding to each of these units as posted by the sellers. For further details on the trading screen's presentation of information and layout, please refer to the screenshots in the subject instructions in the URL above..

Subjects, questionnaire, and subject payments

At the end of each session, the subjects are asked to complete a short survey. The subjects are mostly undergraduate students at a large urban university in the United States. They are nearly equally divided among class ranks, with twenty-five percent freshmen, thirty-one percent sophomores, twenty percent juniors, and twenty-two percent seniors. Masters students make up the remaining two percent of the sample. Approximately seventy-five percent of the subjects have previous experience in an experiment; fifty-nine percent are female; fifty-one percent are African-American; twenty percent are white; and fifteen percent are foreign born. The remaining fourteen percent are either Asian-American, Hispanic-American, or mixed race. The average age in our sample is twenty years old; the minimum age is eighteen years old; and the maximum age is twenty-four years old. Additional descriptive information about the subjects is reported in Table A.1 in the Online Supplement.

After completing the survey, subjects are paid their cumulative earnings for all thirty trading periods, using the conversion rate (one ED = $0.07) announced at the beginning of a session. We pay subjects in a manner that prevents other subjects from observing their payment. Table A.2 in the Online Supplement reports the minimum, average, and maximum earnings for each of the four baseline treatments. The average earnings by treatment are DATS ($26.99), DATB ($27.66), POTS ($20.12), and POTB ($20.79). Completing a session takes approximately two hours; thus, average earnings exceed ten U.S. dollars per hour, which is a favorable hourly wage rate for student subjects. Subjects also receive a five-dollar show-up fee over and above their cumulative earnings from participation in a session.

Analysis of the data from the baseline treatments

In the initial trading periods of a session, there is a lot of "noise" in the transaction price series generated by a market, which reflects the price discovery process of the market institution. However, this noise fades away by the 15th trading period. Once transaction prices converge to an equilibrium (not necessarily a competitive one), subsequent transaction prices vary about the equilibrium price. In short, our laboratory market experiments produce time series data. Although we use a variety of nonparametric and parametric procedures for hypothesis testing, all of them presuppose that the data across trading periods are independent. We conclude from a variety of statistical tests, which are discussed in greater detail in the Online Supplement, that it is appropriate to treat the time series generated by the last fifteen trading periods of our laboratory market experiments as the sum of a constant and an independent white-noise process. We interpret the expected value of the constant to be an equilibrium price. Following the convention in the existing literature, we use the average buyer prices from each of the last fifteen trading periods for hypothesis testing.[8]

Before turning to the discussion of our empirical results, we proceed below by describing our data analysis strategy.

Data analysis strategy

We report results from several distinct tests to ensure that our reported conclusions are robust to alternative statistical procedures. In addition to the Student's t-test with unequal variances, we report the results of the Mann-Whitney-Wilcoxon (MWW) test and the Kolmogorov-Smirnov (KS) test.[9] The properties of these three nonparametric tests are described in the Online Supplement in greater detail. We proceed below by briefly describing a regression-based test of the null hypotheses that are the focus of this research; then, we discuss our empirical results.

We estimate the following RE model:

$$\bar{\mathrm{P}}_{\mathrm{it}} = \beta_0 + \beta_1 \mathrm{T} + \gamma_t + \varepsilon_{it}. \tag{1}$$

The dependent variable in (1) is the average buyer transaction price for trading period t (= 16,…,30) of treatment i (= 1,…, 4 are the four treatments cells of the 2×2 design). In the test of the LSE hypothesis, T=1 when the liability to pay the tax is on the buyer, and zero otherwise. In the test of the IMI hypothesis, T=1 when the market institution is PO, and zero otherwise. The random effect is designated by γ_t for the t^{th} trading period; and ε_{it} is a stochastic error term. We cannot make valid inferences using robust standard errors because Skewness-Kurtosis test rejects the null hypothesis that the residuals from the four baseline RE regressions are normally distributed. Therefore, we use bootstrap standard errors to improve inference. The bootstrap distribution provides an almost exact approximation of the empirical distribution of the sample. In addition, Cameron et al. (2008) show that bootstrap clustered standard errors improve inference when there are a small number of clusters. Accordingly, we report p-values computed with bootstrap clustered (at the market level) standard errors, using five thousand replications.

Results from the four baseline treatments

Now, we are ready to discuss our empirical results. We investigate two questions in this research. Is tax incidence independent of the assignment of the liability to pay tax for a given market institution? Is tax incidence independent of the market institution, keeping the assignment of the liability to pay the tax the same? The former question is the LSE hypothesis, and the latter is the IMI hypothesis.

We begin with a preliminary investigation of the data. Figure 1 consists of a set of four paired bar graphs illustrating the difference in the average buyer prices produced by the last fifteen trading periods from the four baseline comparisons of our two-by-two design.[10] The differences in the relative heights of the paired bars reflects the differences in the average buyer prices of these pairwise comparisons. The results of the Student's t-tests of the null hypothesis of no difference in the averages of the buyer prices and the associated p-values for each comparison are reported in Table 1. Since the Student's t-test has the virtue of familiarity, we use this test for our preliminary look at the data.

The LSE hypothesis predicts that tax-inclusive prices should be the same whether a tax is placed on the buyer or the seller. We test this hypothesis for the DA market institution by comparing the averages of the buyer transaction prices produced by the DATB and DATS markets. As reflected in the first pair of bar graphs of Figure 1, the difference in the averages of the buyer transaction prices produced by these two markets is equal to 0.61 ED. The Student's t-test of the null hypothesis of no difference in the averages of the buyer prices -- reported in panel B1 of Table 1 -- is equal to 5.12 (p-value = 0.000). Thus, the data reject the LSE hypothesis at conventional levels of statistical significance. This finding is rather surprising given the previous discussion of the competitive equilibrium price-convergence properties of DA markets.

We also test the LSE hypothesis for the PO market institution. As reflected in the relative heights of the second pair of bar graphs in Figure 1, the difference in the averages of the buyer transaction prices produced by the POTB and POTS markets is equal to 0.79 ED. The Student's t-test of no difference in the averages of the buyer transaction prices is equal to 5.85 (p-value = 0.000), which is reported in panel B2 of Table 1. This test rejects the LSE hypothesis for PO markets at conventional levels of statistical significance. This finding is less surprising given the previous discussion of the price-convergence properties of PO markets.

Now, we test the IMI hypothesis. Given the differences in the price-convergence properties of DA and PO markets, we expect that that the data will reject this hypothesis. We begin by testing the IMI hypothesis by comparing the averages of the buyer transaction prices produced by the DATB and POTB markets. As evident from the relative heights of the third pair of bars in Figure 1, the difference for this comparison is equal to 2.15 ED. As reported in panel B3 of Table 1, the Student's t-test of no difference in the averages of the buyer transaction prices is equal to 25.85 (p-value = 0.000). The data reject the IMI hypothesis for this comparison at conventional levels of statistical significance. Finally, the difference in the averages of the buyer transactions prices in the comparison of the POTS and DATS markets is 2.33 ED, which is reflected in the difference in the heights of the fourth pair of bars in Figure 1. As reported in panel B4 of Table 1, the Student's t-test of no difference in the averages in the buyer transaction prices from these two treatments is equal to 14.58 (p-value = 0.000). The data reject IMI for this comparison at conventional levels of significance, as well.

The Student's t-test assumes that buyer prices are normally distributed. If the buyer prices are independent draws from a normal distribution, then this assumption would be justified. However, Skewness-Kurtosis test rejects this hypothesis for the data produced by our four baseline treatments. Therefore, inferences based on the Student's t-test should be interpreted with caution, suggesting the importance of the nonparametric tests that we report next.

Tests of the liability side equivalence hypothesis

As before, we begin by examining the LSE hypothesis for DA markets by comparing the average buyer transaction prices produced by the DATB and DATS treatments. The results of the MWW, KS, and RE model tests of the null hypothesis of no price difference are reported in panel B1 of Table 1. The MWW test is equal to -5.800 and is statistically significant at conventional levels. The data reject the null hypothesis that these two price series come from the same population. This finding is inconsistent with the prediction of the LSE hypothesis. The KS and RE-model tests also reject the LSE hypothesis for DA markets at conventional levels of statistical significance.

The results of our tests of the LSE hypothesis for PO markets are reported in panel B2 of Table 1. The MWW test is equal to -4.685 and is statistically significant at conventional levels. Again, the data reject the null hypothesis that the POTB and POTS average buyer transaction prices come from the same population. This finding is inconsistent with the prediction of the LSE hypothesis. The KS and RE-model tests also reject the LSE hypothesis for PO markets at conventional levels of statistical significance.

Tests of the independence of market institutions hypothesis

Now, we turn to our second question: Is the incidence of a tax independent of the type of market institution, holding the assignment of the liability to pay tax the same. Initially, we test the IMI hypothesis by comparing the averages of the buyer transaction prices produced by the DATB and POTB markets from the final fifteen trading periods. The tests of this hypothesis are reported in panel B3 of Table 1. The MWW test of -43.511 rejects the null hypothesis that these two price-series come from the same population at conventional levels of statistical significance. This finding is inconsistent with the prediction of the IMI hypothesis. The KS and RE-model tests also reject the IMI hypothesis for these two market institutions at conventional levels of statistical significance.

Finally, we test the IMI hypothesis by comparing the average buyer transaction prices from the DATS and POTS markets. The results of the tests for this hypothesis are reported in panel B4 of Table 1. The MWW test, which is equal to -9.316, rejects the null hypothesis that these two price series come from the same population at conventional levels of statistical significance. This finding is inconsistent with the prediction of the IMI hypothesis. The KS and RE-model tests also reject the IMI hypothesis for these two market institutions at conventional levels of statistical significance.

In sum, the data generated by our baseline treatments soundly reject the LSE hypothesis for both the DA and PO market institutions using a variety of tests. Similarly, the data rejects the IMI hypothesis for the DATB and POTB treatments as well as the DATS and POTS treatments using a variety of tests.

Are the observed differences economically meaningful?

We report ample evidence rejecting the LSE and IMI hypotheses. Although the price differences discussed above are statistically significant at conventional levels using a variety of tests, it is also important to examine whether these differences are economically significant.

We use several measures of economic significance. We compare the proportions of the total tax revenue paid by buyers and sellers in the four baseline treatments and, in Table 2, we report the average quantities, average excess burdens, excess burden as a proportion of tax revenue, and excess burden as a proportion of participant earnings using the data generated by the last fifteen trading periods for each of the four baseline comparisons.

Since we use a symmetric design, the theory of tax incidence predicts that the tax burden will be equally shared between buyers and sellers. We use the average value of the buyer transaction prices from the last fifteen trading periods from all four markets of the DATB treatment to identify an equilibrium price. As discussed in greater detail below, the average quantity in the DATB treatment is statistically indistinguishable from fifteen units at conventional levels of significance, and fifteen units is the post-tax equilibrium quantity predicted by the theory for the induced demand and supply schedules. This justifies using the average buyer transaction price from the DATB treatment for hypothesis testing. We then use this equilibrium price to determine the incidence of the tax in each of our treatments. In the DATB treatment, fifty percent of the tax burden is borne by buyers and fifty percent by sellers in the DATB treatment. In the DATS markets, 55.1 percent of the tax revenue of 180.4 ED is borne by buyers and 44.9 percent by sellers. In contrast, 67.9 percent of the tax revenue of 133.6 ED is borne by buyers and 32.1 percent by sellers in the POTB markets. The burden of the tax is even further shifted onto buyers in the POTS markets. In this case, 74.5 percent of the tax revenue of 124.8 ED is borne by buyers and 25.5 percent by sellers. Table A.3 of the Online Supplement reports average buyer prices and tax burdens as a percent of total tax revenues for the baseline treatments. Clearly, these differences in tax incidence among the four treatments are economically meaningful.

Now, we turn to the analysis of excess burdens. In columns one through four and rows one through four of Table 2, we report whether the average quantity and average excess burden from the last fifteen trading periods are equal to the theoretical values of fifteen units and sixty ED, respectively, for each of the four baseline treatments. The average quantities reported in Table 2 are consistent with the observed differences in tax shifting documented in Table 1. The average quantity in the DATB treatment is statistically indistinguishable from fifteen units, which is the post-tax equilibrium quantity predicted by the theory, at conventional levels of significance. Furthermore, the average excess burdens from the DATB treatments are statistically indistinguishable from sixty ED at conventional levels of significance, and sixty ED is the predicted value for a competitive equilibrium. In the DATS treatment, the average quantity is 15.03, which is statistically indistinguishable from the competitive equilibrium quantity of fifteen at conventional levels of significance. However, the average excess burden is 63.85 ED, which is statistically different from the theoretical value at conventional values of significance. In contrast, the average quantities and excess burdens are strikingly different from the values predicted by competitive equilibrium in the POTB and POTS treatments. The average quantity and average excess burden are statistically different from fifteen units and sixty ED, respectively, at conventional levels of significance. Table 2 also shows that there are stark differences in excess burdens as a share of tax revenue and as a share of participant earnings among these treatments.

In short, the data generated by the DATB treatment are consistent with the theory of tax incidence in symmetric markets as this treatment replicates the theoretical values of equilibrium quantity and excess burden. Using this treatment as the benchmark, we find substantial differences in the average prices and quantities as well as the excess burdens as a share of participant earnings among the other three baseline treatments.

In sum, contrary to the predictions of the LSE hypothesis, we report evidence that tax incidence is not independent of the assignment of the liability to pay tax nor is it independent of the market institution. We also find that these differences are economically meaningful for all but the DATB market institution.

Robustness tests using alternative experimental designs

We now ask whether our conclusions about the LSE and IMI hypotheses are robust to changing two features of our experimental design. One feature is the information provided to subjects in PO markets. The second feature is the across-sessions treatment design in which market institution and liability to pay tax both vary across sessions.

We now provide sellers with posted prices and quantity limits (henceforth price-quantity information) of competing sellers in the PO markets. We also control for possible session effects by varying the liability to pay tax across markets within sessions of both DA and PO markets. This creates eight additional treatment cells.

We begin by changing the price-quantity information provided to sellers in PO markets. In contrast to the low-information, baseline PO treatments in which sellers cannot observe the posted price-quantity information of other sellers, we now allow sellers to observe the price-quantity information posted by the other sellers in PO markets. We refer to the latter as the high information treatment. As in the baseline treatment, we allow four POTB markets to trade simultaneously in one session and four POTS markets to trade simultaneously in another session. The result of this pairwise comparison is reported in column two of the uppermost panel of Table 3. As in the baseline treatment, all four tests reject the null hypothesis of no price difference in the comparison of POTB and POTS markets. Therefore, we reject the LSE hypothesis at conventional levels of significance for PO markets when sellers can observe the price-quantity information of other sellers in the market.

Next, we control for potential session effects by allowing two POTB and two POTS markets to trade simultaneously in the same session. We conduct this treatment providing low information as in the baseline treatments and high information as in the alternative treatment that we just discussed in the previous paragraph. The results of these treatments are reported in column two of the middle and lowermost panels, respectively, of Table 3. Again, we reject the LSE hypothesis at conventional levels of statistical significance with all the tests, with two exceptions. The MWW and the RE-model tests fail to reject the LSE hypothesis when the PO treatments are run simultaneously in the same session with low information.

We also report measures of economic significance for these alternative treatments in Table 2 and Table A.3 of the Online Supplement. The average quantities and average excess burdens created by the markets in the alternative PO treatments are statistically significantly different from the theoretical values of fifteen units and sixty ED, respectively, at conventional levels of significance.

Turning to the DA markets, we control for session effects by allowing two DATB and two DATS markets to trade simultaneously during the same session.[11] The results of this experimental design are reported in column three of the lowermost panel of Table 3. Both the MWW and Student's t-test reject the null hypothesis of no price difference, but the KS and RE model tests fail to reject the LSE hypothesis. In these two cases, the average quantities and average excess burdens are not statistically distinguishable from the values predicted by theory at conventional levels of statistical significance.

Now we turn to the alternative treatments for the IMI hypothesis, the results of which are reported in Table 4. There are six additional comparisons for this hypothesis. The tests reported in the top, middle, and bottom panels show that our baseline conclusions rejecting the IMI hypothesis are robust to the two alternative experimental designs considered here.

Comparison to previous literature and external validity

Our conclusions differ from those of Borck et al. (2002) who fail to reject the hypothesis of liability side equivalence (LSE) with data from their posted offer (PO) markets. The discrepancy between their conclusions and ours may reflect differences in our respective experimental designs and protocols. However, it may also reflect differences in our respective statistical procedures. As we have explained, our conclusions are based on several nonparametric and parametric tests. Their conclusions are based solely on a test for statistical significance of a liability-treatment dummy variable from a random-effects, panel-data regression model using transaction prices from the last six rounds of their experiment (Borck et al. 2002, pg. 680).

They generously provided us with the data from their PO treatments. Using transaction prices from the last six rounds of these data, we replicated the summary statistics reported in Table 1 of their paper and their significance test which fails to reject the LSE hypothesis at conventional levels of statistical significance. However, we are not able to replicate the exact magnitude of the RE estimator reported in their paper. We begin our analysis of their statistical procedures by computing the Skewness-Kurtosis test for normality of the distribution of the residuals from the RE model that we estimated using their data. This test rejects normality of the error distribution at a p-value of 0.06, which suggests inference from a Student's t-test of the estimated coefficient may not be valid and should be interpreted with caution. Additionally, we estimated bootstrapped clustered (at the market level) standard errors of an RE model using their data. The resulting p-value is equal to 0.063. The Kolmogorov-Smirnov test rejects LSE (p-value = 0.016) as does the Student t-test with unequal variances (p-value = 0.045). The Mann-Whitney-Wilcoxon test weakly rejects LSE with their data (p-value = 0.085). Furthermore, if the LSE hypothesis is empirically valid then it should be equally likely that the average observed price with liability on the seller would be greater or less than the average observed price with liability on the buyer. Inspection of the data reported in Table 1 of the Borck et al. paper reveals that the average buyer transaction prices with buyer liability (POTB) are greater than the average prices with seller liability (POTS) in every round of the final six periods which they use for hypothesis testing. The nonparametric sign test rejects the LSE hypothesis (p-value = 0.016). We conclude that the Borck et al. PO data reject LSE at conventional levels of statistical significance.

We conclude this section with a brief comment on the external validity of our results. We report evidence of greater tax shifting in the POTS treatments, which is consistent with econometric evidence of tax shifting in the analogous field institution. Hanson and Sullivan (2009) use the occasion of a one dollar per pack increase in the cigarette excise tax in Wisconsin to estimate the incidence of a tax on cigarettes. They report evidence that the tax increase is over shifted: the price per pack increases by $1.08 to $1.17 as result of a one dollar increase in the cigarette excise tax. Harris (1987) and Keeler et al. (1996) also report similar evidence of over shifting of cigarette excise taxes, using state level data for multiple states. Finally, Besley and Rosen (1999) examine the incidence of state sales taxes. They report evidence of full shifting of the tax for many commodities and over shifting for more than half of them. It is interesting to find field evidence that is broadly consistent with the evidence from our laboratory POTS market treatments.

Concluding remarks

We analyze data from an experiment designed to examine two important questions regarding tax incidence. Is tax incidence independent of the assignment of the liability to pay tax in either a double auction market or a posted offer market? Is tax incidence independent of the market institution existing in the taxed market? We report robust evidence that the assignment of the liability to pay tax, holding the market institution the same, and the type of market institution, holding the assignment of the liability to pay tax the same, both have a statistically significant effect on the incidence of a tax in the two market institutions examined in this study.

We find that these differences in incidence are economically meaningful in almost all comparisons. We report robust conclusions that hypotheses of (a) liability side equivalence and (b) independence of market institutions can be rejected with data from our experiments.

Given the prominence of the LSE hypothesis in the public finance literature and the practical importance of tax incidence for tax policy design, we believe that our central findings that tax incidence depends on both assignment of liability to pay and on the existing market institution in the taxed market merit further study.

References

Besley, Timothy, and Harvey S. Rosen. 1999. Sales taxes and prices: An empirical analysis. *National Tax Journal* 52 (2): 157–178.

Borck, Rainald, Dirk Engelmann, Wieland Müller, and Hans-Theo Normann. 2002. Tax liability-side equivalence in experimental posted-offer markets. *Southern Economic Journal* 68 (3): 672-682.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics* 90 (3): 414-427.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. Salience and taxation: Theory and evidence. *The American Economic Review* 99 (4): 1145-1177.

Finkelstein, Amy N. 2009. EZ-Tax: Tax salience and tax rates. *The Quarterly Journal of Economics* 124 (3): 969-1010.

Hanson, Andrew, and Ryan Sullivan. 2009. The incidence of tobacco taxation: Evidence from geographic micro-level data. *National Tax Journal* 62 (4): 677-698.

Harris, Jeffrey E. 1987. The 1983 increase in the federal cigarette excise tax. In *Tax Policy and the Economy* (Volume 1), Lawrence H. Summers, ed., 87-111. Cambridge, MA: MIT Press.

Hong, James T., and Charles R. Plott. 1982. Rate filing policies for inland water transportation: An experimental approach. *The Bell Journal of Economics* 13 (1): 1-19.

Kachelmeier, Steven J., Stephen T. Limberg, and Michael S. Schadewald. 1994. Experimental evidence of market reactions to new consumption taxes. *Contemporary Accounting Research* 10 (2): 505-545.

Keeler, Theodore E., Teh-Wei Hu, Paul G. Barnett, Willard G. Manning, and Hai-Yen Sung. 1996. Do cigarette producers price-discriminate by state? An empirical analysis of local cigarette pricing and taxation. *Journal of Health Economics* 15 (4): 499-512.

Kerschbamer, Rudolf, and Georg Kirchsteiger. 2000. Theoretically robust but empirically invalid? An experimental investigation into tax equivalence. *Economic Theory* 16 (3): 719-734.

Ketcham, Jon, Vernon L. Smith, and Arlington W. Williams. 1984. A comparison of posted offer and double auction pricing institutions. *The Review of Economic Studies* 51 (4): 595-614.

Morone, Andrea, and Francesco Nemore. 2015. Tax salience: An experimental investigation. Munich Personal RePEc Paper, Munich, Germany.

Plott, Charles R., and Vernon L. Smith. 1978. An experimental examination of two exchange institutions. *The Review of Economic Studies* 45 (1): 113-153.

Riedl, Amo, and Jean-Robert Tyran. 2005. Tax liability side equivalence in gift-exchange labor markets. *Journal of Public Economics* 89 (11): 2369-2382.

Ruffle, Bradley J. 2005. Tax and subsidy incidence equivalence theories: Experimental evidence from competitive markets. *Journal of Public Economics* 89 (8): 1519-1542.

Saez, Emmanuel, Manos Matsaganis, and Panos Tsakloglou. 2012. Earnings determination and taxes: Evidence from a cohort-based payroll tax reform in Greece. *The Quarterly Journal of Economics* 127 (1): 493-533.

Sausgruber, Rupert, and Jean-Robert Tyran. 2005. Testing the Mill hypothesis of fiscal illusion. *Public Choice* 122 (1): 39-68.

Smith, Vernon L. 1976a. Experimental economics: Induced value theory. In *The American Economic Review, Papers and Proceedings of the Eighty-Eighth Annual Meeting*, 274-279. American Economic Association. Nashville, TN.

Smith, Vernon L. 1976b. Bidding and auctioning institutions: Experimental results. In *Bidding and Auctioning for Procurement and Allocation*, Yakov Amihud, ed., 43-64. New York, NY: New York University Press.

Smith, Vernon L. 1982a. Microeconomic systems as an experimental science. *The American Economic Review* 72 (5): 923-955.

Smith, Vernon L. 1982b. Reflections on some experimental market mechanisms for classical environments. In *Choice Models for Buyer Behavior: A Research Annual,* Leigh McAlister, ed., 13-47. Greenwich, CT: JAI Press, Inc.

Smith, Vernon L. 1982c. Markets as economizers of information: Experimental examination of the Hayek hypothesis. *Economic Inquiry* 20 (2): 165-179.

Smith, Vernon L., and Arlington W. Williams. 1982. The effects of rent asymmetries in experimental auction markets. *Journal of Economic Behavior and Organization* 3 (1): 99-116.

Smith, Vernon L., and Arlington W. Williams. 1983. An experimental comparison of alternative rules for competitive market exchange. In *Auctions, Bidding and Contracting: Uses and Theory*, Richard Engelbrecht-Wiggans, Martin Shubik, and Robert M. Stark, ed., 307-334. New York: New York University Press.

Smith, Vernon L., Arlington W. Williams, W. Kenneth Bratton, and Michael G. Vannoni. 1982. Competitive market institutions: Double auctions versus sealed bid offer auctions. *The American Economic Review* 72 (1): 58-77.

Weyl, E. Glen, and Michal Fabinger. 2013. Pass-through as an economic tool: Principles of incidence under imperfect competition. *Journal of Political Economy* 121 (3): 528-583.

Williams, Arlington W. 1980. Computerized double-auction markets: Some initial experimental results. *The Journal of Business* 53 (3): 235-258.

Author Biographies

James C. Cox is the Noah Langdale Jr. Chair in Economics, Georgia Research Alliance Eminent Scholar, and Director of the Experimental Economics Center at Georgia State University. He received his doctorate in economics from Harvard University. His current research includes behavioral experiments and theoretical modeling of: trust, reciprocity, and altruism; trust and trustworthiness of immigrants and native-born Americans; calibration paradoxes for small- and large-stakes risk aversion; incentive compatibility of payoff mechanisms for choice under risk; public goods and common pool resources; isomorphic centipede games and Dutch auctions; decision making in organ transplantation surgery; and clinical decision support systems for healthcare.

Mark Rider is a professor in the Department of Economics at Georgia State University. He received his doctorate in economics from Georgia State University. He has authored scholarly articles on tax evasion and behavioral responses to taxes and worked on tax reforms in a number of developing and transition economies.

Astha Sen is a professor in the Department of Economics at Sonoma State University. She received her doctorate in economics from Georgia State University. She is interested in studying the effectiveness of tax policy reforms in the developed and developing economies. Her ongoing research examines economic and revenue efficiency of adoption of a value-added tax (VAT) at the sub-national level in India.

Table 1. Four baseline tests [a]

Test statistics	Baseline tests of the liability side equivalence hypothesis	
	B1: DATB versus DATS	**B2:** POTB versus POTS
	Number of observations = 120	Number of observations = 120
Student's t-test with unequal variances	-5.116***	-5.852***
Mann-Whitney-Wilcoxon two-sample test	-5.800***	-4.685***
Kolmogorov-Smirnov two-sample test	0.617***	0.433***
Estimated coefficient from RE model (β_T) (with bootstrap standard errors)	-0.610*	-0.790**
Test statistics	Baseline tests of the tax-incidence equivalence of market institutions hypothesis	
	B3: DATB versus POTB	**B4:** DATS versus POTS
	Number of observations = 120	Number of observations = 120
Student's t-test with unequal variances	-25.850***	-14.580***
Mann-Whitney-Wilcoxon two-sample test	-43.511***	-9.316***
Kolmogorov-Smirnov two-sample test	0.983***	0.917***
Estimated coefficient from RE model (β_T) (with bootstrap standard errors)	2.149***	2.329***

[a] * Statistically significant at the ten-percent level; ** statistically significant at the five-percent level; and *** statistically significant at the one-percent level.

Table 2. Average quantities and excess burdens [a]

Treatment	Average quantity (Units)[b]	Average excess burden (ED)[c]	Excess burden as a percent of tax revenue	Excess burden as a percent of participant earnings
1. Double auction, tax on buyer, four markets trading in one session.	15.10	60.82	33.44	21.91
2. Double auction, tax on seller, four markets trading in one session.	15.03	63.85***	35.39	23.65
3. Posted offer, tax on buyer, low information, four markets trading in one session.	11.13***	134.70***	100.82	64.78
4. Posted offer, tax on seller, low information, four markets trading in one session.	10.40***	147.40***	118.11	73.26
5. Double auction, tax on buyer, two markets trading in a mixed session	15.17	61.20	33.62	22.26
6. Double auction, tax on seller, two markets trading in a mixed session	15.17	62.40	34.28	22.93
7. Posted offer, tax on buyer, high information, four markets trading	13.45***	86.25***	53.44	33.34
8. Posted offer, tax on seller, high information, four markets trading	13.53***	88.40***	54.45	34.95
9. Posted offer, tax on buyer, low information, two markets trading in a mixed session with row 10	10.70***	142.00***	110.59	67.38
10. Posted offer, tax on seller, low information, two markets trading in a mixed session with row 9	11.77***	120.70***	85.46	52.34
11. Posted offer, tax on buyer, high information, two markets trading in mixed session with row 12	13.87***	79.7***	47.89	31.16
12. Posted offer, tax on buyer, high information, two markets trading in mixed session with row 11	13.13***	91.40***	58.01	35.72

[a] *** statistically significant at the one-percent level.

[b] Tests of the null hypothesis that the quantity is equal to fifteen units, which is the value predicted by the theory.

[c] Tests of the null hypothesis that the excess burden is equal to sixty ED, which is the value predicted by the theory.

Table 3. Robustness tests of the liability side equivalence hypothesis [a]

Experimental design	POTB vs. POTS	DATB vs. DATS
1. High information - different sessions	Number of observations = 120	- [b]
Mann-Whitney-Wilcoxon two-sample test	4.803***	-
Kolmogorov-Smirnov two-sample test	0.450***	-
Estimated coefficient from RE model (β_T) (with bootstrap standard errors)	0.374***	-
Student's t-test with unequal variances	10.139***	-
2. Low information - same session	Number of observations = 60	- [c]
Mann-Whitney-Wilcoxon two-sample test	1.619	-
Kolmogorov-Smirnov two-sample test	0.333***	-
Estimated coefficient from RE model (β_T) (with bootstrap standard errors)	0.296	-
Student's t-test with unequal variances	5.365***	-
3. High information - same session	Number of observations = 60	Number of observations = 60
Mann-Whitney-Wilcoxon two-sample test	-5.566***	-1.671*
Kolmogorov-Smirnov two-sample test	0.700***	0.300
Estimated coefficient from RE model (β_T) (with bootstrap standard errors)	-0.378***	-0.170
Student's t-test with unequal variances	-7.497***	-2.483**

[a] * Statistically significant at the ten-percent level; ** statistically significant at the five-percent level; and *** statistically significant at the one-percent level.

[b] This is the baseline test of the liability side equivalence hypothesis for DA markets, which is reported in the panel labelled B2 in Table 3. These tests are repeated here for ease of comparison.

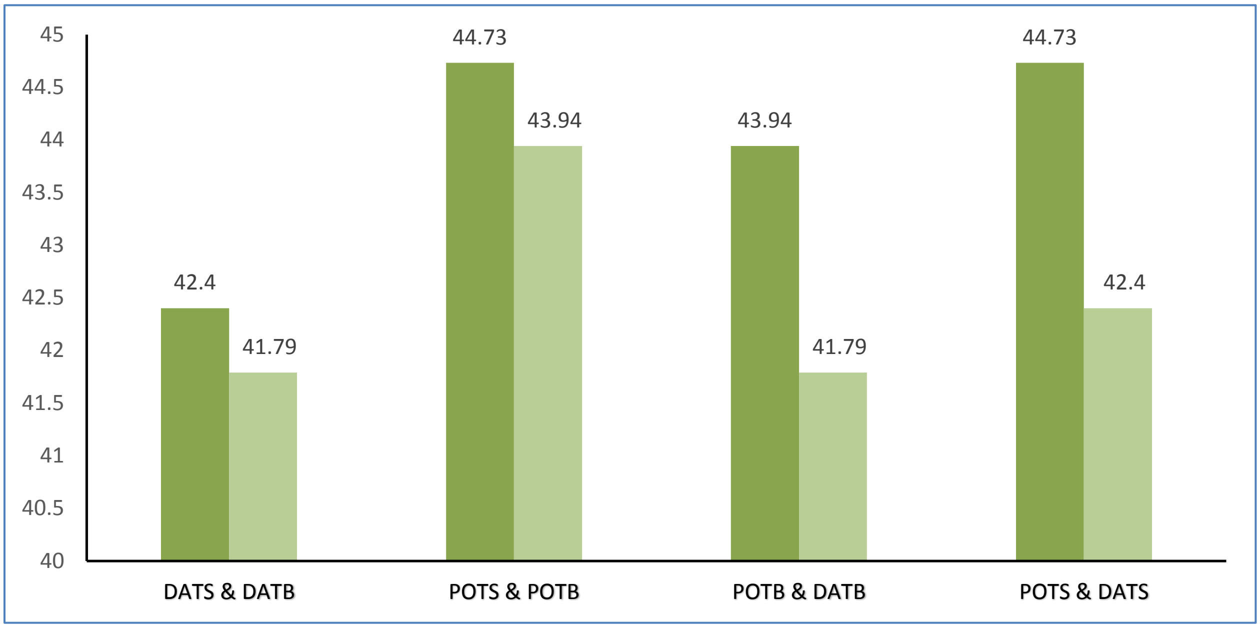
[c] The DA market institution by nature is a high information environment. Therefore, the low information attribute is not applicable here.

Table 4. Robustness tests of the tax-incidence equivalence of market institutions hypothesis [a]

Experimental design	DATB vs. POTB	DATS vs. POTS
1. Posted offer market is high information; DA and PO markets traded in separate sessions	Number of observations = 120	Number of observations = 120
Mann-Whitney-Wilcoxon two-sample test	-9.448***	-6.523***
Kolmogorov-Smirnov two-sample test	1.000***	0.567***
Estimated coefficient from RE model (β_T) (with bootstrap standard errors)	1.941***	0.957**
Student's t-test with unequal variances	-56.057***	-11.434***
2. Posted offer market is low information; DA and PO markets traded in the same session	Number of observations = 60	Number of observations = 60
Mann-Whitney-Wilcoxon two-sample test	-6.654***	-6.653***
Kolmogorov-Smirnov two-sample test	1.000***	1.000***
Estimated coefficient from RE model (β_T) (with bootstrap standard errors)	2.261***	1.795***
Student's t-test with unequal variances	-30.554***	-29.087***
3. Post-offer market is high information; DA and PO markets traded in the same session	Number of observations = 60	Number of observations = 60
Mann-Whitney-Wilcoxon two-sample test	-6.210***	-6.653***
Kolmogorov-Smirnov two-sample test	0.933***	1.000***
Estimated coefficient from RE model (β_T) (with bootstrap standard errors)	1.290***	1.498***
Student's t-test with unequal variances	-16.983***	-38.360***

[a] * Statistically significant at the ten-percent level; ** statistically significant at the five-percent level; and *** statistically significant at the one-percent level.

Figure 1. Pair-wise comparisons of average buyer prices, using the last 15 trading periods of each market session



Notes: DATB = double auction market with tax on buyer.

DATS = double auction market with tax on seller.

POTB = posted offer market with tax on buyer.

POTS = posted offer market with tax on seller.

Acknowledgements

We appreciate helpful comments and suggestions by the editor and three anonymous referees.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The National Science Foundation (grant number SES-0849590) provided partial financial support.

Notes

[1] Taxes could also be shifted backward onto the factors of production.

[2] An exception is Saez et al. (2012) who test the LSE hypothesis using data from change in the assignment of the liability to pay the payroll tax in Greece. Their data is not consistent with LSE.

[3] Using a laboratory experiment, Sausgruber and Tyran (2005) test the "Mill hypothesis" that the burden of an indirect tax may differ from that of a direct tax because the latter is more visible or salient. In a study using a field experiment and observational data on gasoline and liquor sales, Chetty et al. (2009) report evidence that tax salience has an impact on tax incidence. Finkelstein (2009) reports that salience influences the elasticity of demand for tolled roads. Morone and Nemore (2015) examine tax salience in a laboratory setting. They report that in tax-on-seller treatment prices are systematically greater, thus revealing a plausible tax-shifting phenomenon.

[4] See, for example, Ketcham, Smith, and Williams (1984), Plott and Smith (1978), Hong and Plott (1982), Smith (1982b), and Ketcham et al. (1984).

[5] Buyers choosing quantities at the posted prices does not, in itself, mean that buyers must be price takers. Buyers can withhold demand to try to get sellers to post lower prices in subsequent market periods. Buyers may, however, choose to fully reveal demand at posted prices.

[6] In "commodity" double auctions such as ours, announcing or not announcing the end period is unlikely to affect trading. In contrast, in an asset double auction in which there can be price bubbles, this design feature can influence the price determination process.

[7] Chetty et al. (2009) defines "tax salience" to buyers as display of the gross-of-tax price (as we do in our experiment.

[8] Interestingly, if we use all the observations, rather than collapsing the observations into average prices for each trading period, our conclusions remain the same.

[9] Henceforth, we will refer to Student's t-test with unequal variances as simply Student's t-test.

[10] The choice of buyer or seller prices has no bearing on inferences drawn from the data. The difference between the buyer price and the seller price is always equal to the unit tax of twelve ED. Comparisons *across* the four treatments would be the same if done with seller prices.

[11] By its very nature, DA markets provide high information on the price offers of other sellers in real time. Therefore, we cannot vary this feature of the experimental design for DA markets.

Online Supplement to

"Tax Incidence: Do Institutions Matter? An Experimental Study"

By James C. Cox[a], Mark Rider[a], and Astha Sen[b]

[a] Experimental Economics Center and Department of Economics, Andrew Young School of Policy Studies, Georgia State University.

[b] School of Business and Economics, Sonoma State University.

Corresponding author: James C. Cox, office phone (404) 413-0200 and email address jccox@gsu.edu.

Introduction

In this Online Supplement to "Tax Incidence: Do Institution Matter? An Experimental Study," we provide a more thorough review of the literature; a figure illustrating the induced supply and demand schedules (Figure A.1) and a related discussion; a description of the sample and subject payments and related tables (Tables A.1 and A.2); a more elaborate description of the data generating process that motivates our approach to hypothesis testing and a related figure (Figure A.3); a discussion of the information conditions used in laboratory posted offer markets; and a discussion about whether the observed differences among our treatments are economically meaningful and a related table (Table A.3).

Literature review

We are not the first to propose testing the LSE hypothesis in a laboratory setting. Kachelmeier, Limberg and Schedewald (1994), Kerschbamer and Kirchsteiger (2000), Borck et al. (2002), Riedl and Tyran (2005), and Ruffle (2005) test the LSE hypothesis in laboratory settings using a variety of tax types and institutions. Kerschbamer and Kirchsteiger (2000) test the LSE hypothesis by imposing a tax on either the proposer or the responder in a one-stage ultimatum bargaining game. They contend that the LSE hypothesis should hold in a tax version of the ultimatum game as long as the fairness norm threshold below which offers are rejected remains unaffected by a change in liability to pay the tax. They reject the LSE hypothesis with ultimatum game data at conventional levels of significance.

Riedl and Tyran (2005) test the LSE hypothesis, using a gift-exchange game, with a PO market for labor. They contend that LSE should hold in gift-exchange games when gross wages adjust to changes in taxes and market participants are exclusively concerned with net-of-tax profits and wages. Riedl and Tyran do not reject the LSE hypothesis at conventional levels of significance.

Kachelmeier, Limberg, and Schedewald (1994) use an experimental design with two interrelated DA markets, three types of agents (customers, retailers, and wholesalers), and three tax instruments (an ad valorem tax levied on customers, a turnover tax levied on retailers, and a value-added tax levied on wholesalers and retailers). In their experiment, there are ten periods without a tax followed by ten periods with a tax. They do not reject the LSE hypothesis at conventional levels of statistical significance.

Borck et al. (2002) test the LSE hypothesis in PO markets, using a unit tax. In their experiment, there is no tax in the first twelve periods of a session followed by a unit tax in periods thirteen through twenty-four. They conduct ten sessions with PO markets and two sessions with DA markets. They do not reject the LSE hypothesis for either market institution. However, they do not report whether tax incidence differs between the two market institutions, which is one of our central questions.

Ruffle (2005) tests the LSE hypothesis for taxes and subsidies. He uses a pit market where at least eight pairs of buyers and sellers participate in all experiments.[11] There is no tax or subsidy in the first eight periods of a session. In periods nine to nineteen there is a tax or subsidy on either the buyers or the sellers. Ruffle's tests do not reject the LSE hypothesis for either subsidies or taxes.

Our experimental design differs from much of the existing literature in a number of important ways. First, our induced supply and demand schedules are symmetric both at the aggregate market level and the individual agent level. As Weyl and Fabinger (2013) show, LSE is independent of the market structure – perfectly or imperfectly competitive – when markets are symmetric. Therefore, we use a symmetric design to test the LSE hypothesis in two prominent market institutions. Second, we introduce a unit tax in the initial trading period and maintain that tax throughout all thirty trading periods. Third, we present the buyers and sellers with information that makes the tax salient (see section 4.3 in the printed text). Fourth, in addition to testing the LSE hypothesis for two market institutions, we also pose and attempt to answer a new hypothesis: is tax incidence independent of the market institution. Fifth, we calculate the excess burden due to the unit tax for each of the four comparisons that result from our two-by-two design to gauge whether the observed differences in tax shifting are economically meaningful. Sixth, we use a variety of statistical tests and conduct a variety of robustness checks.

Induced Demand and Supply Schedules

Following Smith (1976a and 1982a), we induce stationary demand and supply functions for a fictitious commodity. Figure A.1 of this Online Supplement illustrates the induced demand and supply schedules used in every treatment reported in this study. More specifically, there are five sellers and five buyers in any given market. Each of the five sellers is assigned a cost of twenty-three experimental dollars (ED) for unit 1, twenty-six ED for unit 2, twenty-nine ED for unit 3, thirty-two ED for unit 4, thirty-five ED for unit 7, and thirty-eight ED for unit 6. These marginal costs identify the induced individual supply schedules used in the experiment. Each of the five buyers is assigned a value of fifty ED for unit 1, forty-seven ED for unit 2, forty-four ED for unit 3, forty-one ED for unit 4, thirty-eight ED for unit 5, and thirty-five ED for unit 6.

Information condition in laboratory posted offer markets

Our choice of the baseline information conditions for PO markets differs from many PO market experiments reported in the existing literature. We proceed below with a discussion of this experimental design feature.

In the baseline treatments of the PO market, a seller's trading screen only lists the offers posted by oneself. Sellers are not able to see the offers posted by other sellers in the market. We believe that this feature of the treatment best reflects the field institution that we are trying to replicate in the laboratory. In retail markets, for example, the PO market institution *itself* does not provide information on the prices posted by other sellers in the way that (bid and) ask prices are publicly provided to all sellers (and buyers) in DA markets. While it is true that retail stores can conduct price surveillance and market studies, conducting such studies makes the point that the market institution itself does not provide pricing information to other sellers in the market.

The original question posed in this article is whether tax incidence varies across market institutions given the information that is provided *by the institutions themselves*. This focus on the information-generating properties of a market institution reflects the view articulated by Smith (1982c). The DA market institution produces public information on bid prices, ask prices, and transaction prices. This is a feature of the market institution, itself, both in the laboratory and in organized exchanges such as the New York Stock Exchange. In contrast, the PO market institution, itself, does not provide all of this information. With the PO market institution, information on competitors' prices would have to come from some (possibly costly) activity − such as market surveillance − that is distinct from own-price posting and exchange in the market. The literature has been unclear on this point. Early PO market experiments were run in classrooms using a blackboard and chalk (Williams 1973); thus, there was no practical alternative to displaying all posted prices to all subjects, regardless of whether they were buyers or sellers.

When it subsequently became possible to run computerized PO experiments, authors recognized that there was a choice to be made about whether to exhibit seller's posted prices to their competitors (Ketchum, Smith, and Williams 1984, p. 599). For reasons that are not clear in the literature, a common procedure in many computerized PO experiments is to reveal competing sellers' prices, even though this is not an inherent feature of the PO market institution. As one of our robustness checks, which are discussed in section 5 of the printed version of the paper, we also report the results of PO treatments in which sellers observe the posted price-quantity pairs of other sellers. This change in the experimental design has no effect on our conclusions. In sum, the present paper rejects the LSE hypothesis in PO markets; this finding is robust to the two alternative ways of implementing the PO market institution discussed above in a symmetric induced valuation environment.

Subjects, questionnaire, and subject payments

At the end of each session, the subjects are asked to complete a short survey. Descriptive information about the subjects is reported in Table A.1 of this Online Supplement. The subjects are mostly undergraduate students at a large urban university in the United States. They are nearly equally divided among class ranks, with twenty-five percent freshmen, thirty-one percent sophomores, twenty percent juniors, and twenty-two percent seniors. Masters students make up the remaining two percent of the sample. Approximately seventy-five percent of the subjects have previous experience in an experiment; fifty-nine percent are female; fifty-one percent are African-American; twenty percent are white; and fifteen percent are foreign born. The remaining fourteen percent are either Asian-American, Hispanic-American, or mixed race. The average age in our sample is twenty years old; the minimum age is eighteen years old; and the maximum age is twenty-four years old.

After completing the survey, subjects are paid their cumulative earnings for all thirty trading periods, using the conversion rate (one ED = $0.07) announced at the beginning of a session. We pay subjects in a manner that prevents other subjects from observing their payment. Table A.2 in the Online Supplement reports the minimum, average, and maximum earnings for each of the four baseline treatments. The average earnings by treatment are DATS ($26.99), DATB ($27.66), POTS ($20.12), and POTB ($20.79). Completing a session takes approximately two hours; thus, average earnings exceed ten U.S. dollars per hour, which is a favorable hourly wage rate for student subjects. Subjects also receive a five-dollar show-up fee over and above their cumulative earnings from participation in a session.

Data generating process: what is an observation?

The properties of the data produced by our laboratory experiments should inform our approach to hypothesis testing. Our laboratory market experiments produce time series data. There is a lot of "noise" in the price series in the initial trading periods in a market, which reflects the price discovery process of the market institution. As we show, however, this noise fades away by the 15th trading period. Once transactions prices have converged to an equilibrium (not necessarily a competitive one) subsequent transaction prices vary about this observed equilibrium price. In other words, the data generating process of our market experiments can be characterized as a moving average (MA) process. To make matters more concrete, a qth order MA process, denoted MA(q), is characterized as:

$$P_t = P_a + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \ldots + \theta_q\varepsilon_{t-q}. \qquad (1)$$

P_t is the observed transaction price for the tth trade (after the 15th trading period); P_a is the average of all observed transaction prices preceding the tth trade; and ε_t is assumed to be a white-noise process or a sequence $\{\varepsilon_t\}$ whose elements have mean zero and finite variance σ^2.

Economic theory does not provide much guidance on the statistical properties of ε_t. In fact, an equilibrium price is not generally thought of as a random variable in the theoretical literature. If, however, a price series produced by our laboratory markets converges to an equilibrium, the resulting random variable is said to be ergodic for the mean which implies that it is a covariance-stationary process or $\theta_s < 1$ for all s (= 1,…,q).[11] Furthermore, if, after a sufficient number of trading periods, these observed transaction prices converge to an equilibrium and, in the words of Paul Samuelson (1993), "vibrate randomly", then ε_t should be an independent white-noise process. This implies that $\theta_s = 0$ for all s (= 1,…, q).[11] Lacking theoretical support for such an assumption, we rely on empirical evidence to guide us.

Figure A.2 in the Online Supplement shows the plots of the autocorrelation functions of the transactions prices for the last fifteen trading periods for each of our four baseline treatments. The estimated autocorrelations in each of the four plots are well within the ninety-five percent confidence intervals, except for the first autocorrelation in the POTS treatment (see Figure A.2d). This autocorrelation is statistically significant at the five-percent level. We also conduct runs tests of serial independence using the last fifteen trading periods from the sixteen baseline markets. These tests fail to reject the null hypothesis of serial independence in fourteen out of the sixteen cases at the ten-percent significance level. Pooling the market data by treatment, the runs tests are equal to -1.26 (p-value = 0.21) for the DATB price series; 1.07 (p-value = 0.29) for the DATS price series; -0.59 (p-value = 0.55) for the POTB price series; and -0.01 (p-value = 1.00) for the POTS price series. Thus, we fail to reject the null hypothesis of serial independence in all four cases.

Based on the totality of the statistical evidence, we believe that it is appropriate to treat the price series from the last fifteen trading periods in our laboratory market experiments as the sum of a constant and an independent white-noise process. We interpret the expected value of the constant to be an equilibrium price. This implies that observed individual transaction prices (from the final fifteen trading periods) can be used in hypothesis tests. Nonetheless, we follow the convention in the existing literature and use the average buyer prices for each of the last fifteen trading periods.[11] However, and this is an important point, the statistical tests employed in this research, and discussed in greater detail below, assume that the data are independent. We cannot reject this hypothesis for our data.

The Student's t-test compares two samples by assuming there is no difference in the sample means. It computes a p-value for the null hypothesis that the sample means are equal by further assuming the samples are independent draws from a normal distribution.[11] This property of the Student's t-test is too restrictive in the current context because Skewness-Kurtosis tests reject the null hypothesis that the data generated by our laboratory market experiments are normally distributed at conventional levels of significance.[11] In contrast, the Mann-Whitney-Wilcoxon (MWW) test compares two samples consisting of independent draws from a population by first ranking all the values from the two samples in ascending order, and then computes a p-value that depends on the discrepancy between the mean ranks of the two samples. In this sense, it is similar to a Student's t-test; however, this nonparametric test requires no assumption about the distribution of the data. The intuition behind the Kolmogorov-Smirnov (KS) test is straightforward. If two samples are independent draws from the same population, the two empirical cumulative distribution functions (CDFs) should be reasonably similar. The KS test computes a p-value for the largest discrepancy between the two CDFs. If the maximum discrepancy is substantial in a statistical sense, the investigator can conclude that there is a high likelihood that the samples come from two different populations. Like the MWW test, the KS test is also nonparametric and requires no assumption about the distribution of the data.

Are the observed differences economically meaningful?

Since we use a symmetric design, the theory of tax incidence predicts that the tax burden will be equally shared between buyers and sellers. We use the average value of the buyer transaction prices from the last fifteen trading periods from all four markets of the DATB treatment to identify an equilibrium price. As discussed in greater detail below, the average quantity in the DATB treatment is statistically indistinguishable from fifteen units at conventional levels of significance, and fifteen units is the post-tax equilibrium quantity predicted by the theory for the induced demand and supply schedules. This justifies using the average buyer transaction price from the DATB treatment for hypothesis testing. We then use this equilibrium price to determine the incidence of the tax in each of our treatments.

Table A.3 of this Online Supplement reports average buyer prices and tax burdens as a percent of total tax revenues. In the DATB treatment, fifty percent of the tax burden is borne by buyers and fifty percent by sellers in the DATB treatment. In the DATS markets, 55.1 percent of the tax revenue of 180.4 ED is borne by buyers and 44.9 percent by sellers. In contrast, 67.9 percent of the tax revenue of 133.6 ED is borne by buyers and 32.1 percent by sellers in the POTB markets. The burden of the tax is even further shifted onto buyers in the POTS markets. In this case, 74.5 percent of the tax revenue of 124.8 ED is borne by buyers and 25.5 percent by sellers. Clearly, these differences in tax incidence among the four treatments are economically meaningful.

References

Borck, Rainald, Dirk Engelmann, Wieland Müller, and Hans-Theo Normann. 2002. Tax liability-side equivalence in experimental posted-offer markets. *Southern Economic Journal* 68 (3): 672-682.

Hamilton, James D. 1994. *Time Series Analysis* (Volume 2). Princeton, NJ: Princeton University Press.

Kachelmeier, Steven J., Stephen T. Limberg, and Michael S. Schadewald. 1994. Experimental evidence of market reactions to new consumption taxes. *Contemporary Accounting Research* 10 (2): 505-545.

Kerschbamer, Rudolf, and Georg Kirchsteiger. 2000. Theoretically robust but empirically invalid? An experimental investigation into tax equivalence. *Economic Theory* 16 (3): 719-734.

Ketcham, Jon, Vernon L. Smith, and Arlington W. Williams. 1984. A comparison of posted offer and double auction pricing institutions. *The Review of Economic Studies* 51 (4): 595-614.

Riedl, Amo, and Jean-Robert Tyran. 2005. Tax liability side equivalence in gift-exchange labor markets. *Journal of Public Economics* 89 (11): 2369-2382.

Ruffle, Bradley J. 2005. Tax and subsidy incidence equivalence theories: Experimental evidence from competitive markets. *Journal of Public Economics* 89 (8): 1519-1542.

Samuelson, Paul A. 1973. Proof that properly discounted present values of assets vibrate randomly. *The Bell Journal of Economics and Management Science* 4 (2): 369-374.

Smith, Vernon L. 1976. Experimental economics: Induced value theory. In *The American Economic Review, Papers and Proceedings of the Eighty-Eighth Annual Meeting*, 274-279. American Economic Association. Nashville, TN.

Smith, Vernon L. 1982a. Microeconomic systems as an experimental science. *The American Economic Review* 72 (5): 923-955.

Smith, Vernon L. 1982b. Markets as economizers of information: Experimental examination of the Hayek hypothesis. *Economic Inquiry* 20 (2): 165-179.

Weyl, E. Glen, and Michal Fabinger. 2013. Pass-through as an economic tool: Principles of incidence under imperfect competition. *Journal of Political Economy* 121 (3): 528-583.

Williams, Fred E. 1973. The effect of market organization on competitive equilibrium: The multi-unit case. *The Review of Economic Studies* 40 (1): 97-113.

Table A.1. Demographic characteristics of the full sample of the baseline treatments

Variable	Baseline Treatments				
	Full	Double Auction		Posted Offer	
		Tax on seller	Tax on buyer	Tax on seller	Tax on buyer
Percent buyers	50	50	50	50	50
Percent freshmen	25	15	30	30	25
Percent sophomore	31	30	30	30	35
Percent juniors	20	28	10	23	20
Percent seniors	22	25	30	18	15
Percent Masters students	2	3	0	0	5
Percent with experience in experiments	76	78	78	75	73
Percent female	59	55	68	63	50
Percent African-American	51	48	43	58	58
Percent Asian-American	4	3	0	5	8
Percent Hispanic-American	2	5	3	0	0
Percent mixed race	8	10	8	10	5
Percent white	20	23	25	20	13
Percent foreign born	15	13	23	8	18
Average age	20	20	20	20	20
Standard deviation of age	2	2	2	1	1
Minimum age	18	18	18	18	18
Maximum age	24	24	24	24	23
Percent business administration majors	28	25	28	43	18
Percent economics majors	6	0	5	3	15
Percent other majors	0	0	0	0	0
Percent with at least 1 economics course	55	53	58	58	53
GPA between 1.25 and 2.74 (percent)	9	8	5	13	13
GPA between 2.75 and 3.24 (percent)	29	48	28	25	18
GPA between 3.25 and 3.74 (percent)	39	30	35	45	45
GPA between 3.75 and 4.0 (percent)	19	10	30	13	23
Not taken courses with grades (percent)	4	5	3	5	3
Number of experimental subjects	160	40	40	40	40

Table A.2. Earnings in U.S. dollars, for the baseline treatments

Earnings	Baseline Treatments				
	Full	Double auction		Posted offer	
		tax on seller	tax on buyer	tax on seller	tax on buyer
Minimum Earnings	$6.75	$14.25	$16.75	$9.25	$6.75
Average Earnings	$23.89	$26.99	$27.66	$20.12	$20.79
Maximum Earnings	$40.75	$38.75	$35.25	$40.75	$38.50

Table A.3. Average buyer prices and tax burden as a percent of tax revenue

Treatment	Average buyer price (ED)	Total tax revenue (ED)	Proportion of the tax revenue paid by the buyer (percent)	Proportion of the tax revenue paid by the seller (percent)
1. Double auction, tax on buyer, four markets trading in one session.	41.8	181.2	50.0	50.0
2. Double auction, tax on seller, four markets trading in one session.	42.4	180.4	55.1	44.9
3. Posted offer, tax on buyer, low information, four markets trading in one session.	43.9	133.6	67.9	32.1
4. Posted offer, tax on seller, low information, four markets trading in one session.	44.7	124.8	74.5	25.5
5. Double auction, tax on buyer, two markets trading in a mixed session with row 6	42.1	182.0	52.9	47.1
6. Double auction, tax on seller, two markets trading in a mixed session with row 5	42.3	182.0	54.3	45.7
7. Posted offer, tax on buyer, high information, four markets trading in one session	43.7	161.4	66.2	33.8
8. Posted offer, tax on seller, high information, four markets trading in one session	43.3	162.4	62.9	37.1
9. Posted offer, tax on buyer, low information, two markets trading in a mixed session with row 10	44.4	128.4	71.7	28.3
10. Posted offer, tax on seller, low information, two markets trading in a mixed session with row 9	44.0	141.2	68.8	31.2
11. Posted offer, tax on buyer, high information, two markets trading in mixed session with row 12	43.4	166.4	63.7	36.3
12. Posted offer, tax on buyer, high information, two markets trading in mixed session with row 11	43.8	157.6	66.8	33.2

Notes: The estimates reported above are based on the data from the last fifteen trading periods of each treatment. The average buyer and seller prices from the DATB treatment are used as the baseline to calculate the incidence of the other three treatments. Figures reported in the table are rounded to first decimal place.

2. Online Figures

Figure A.1. Induced demand and supply schedules

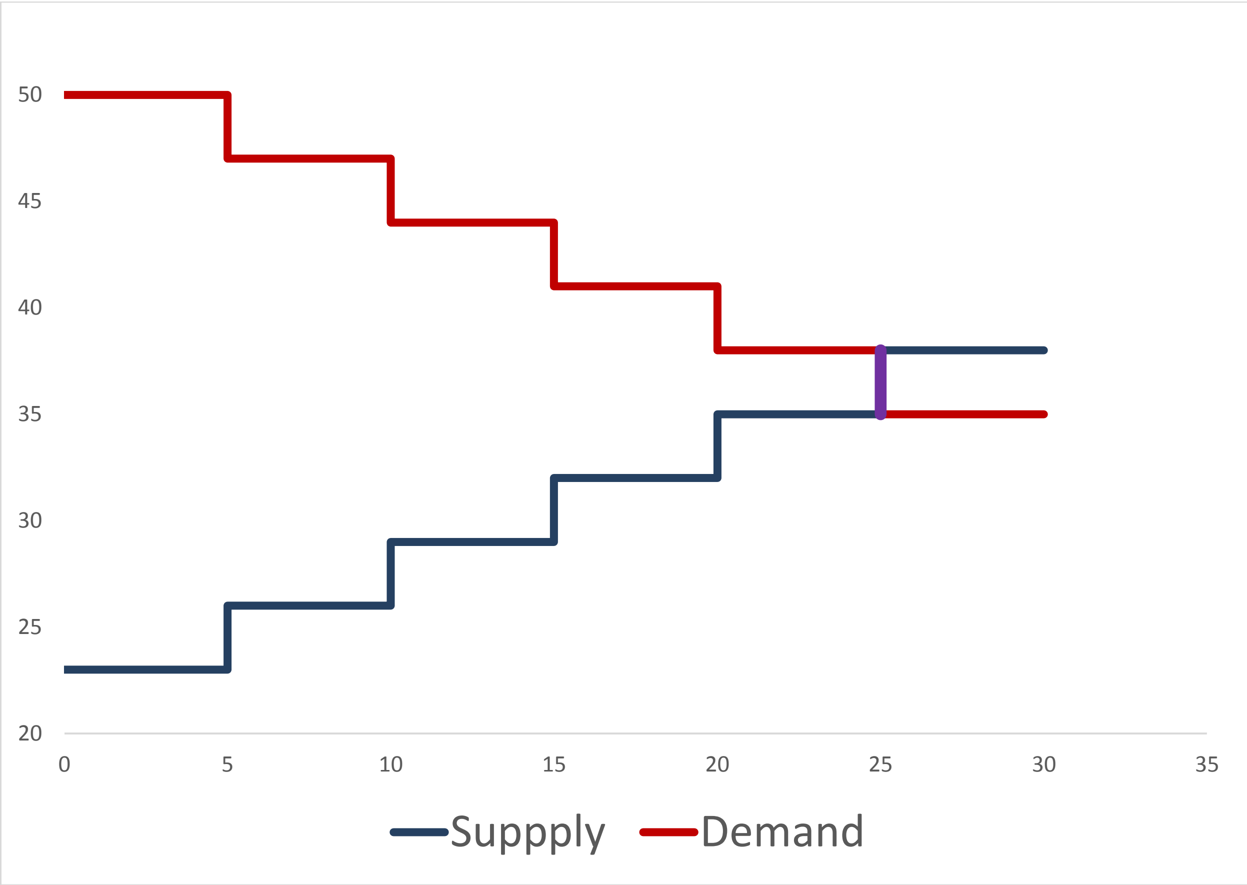
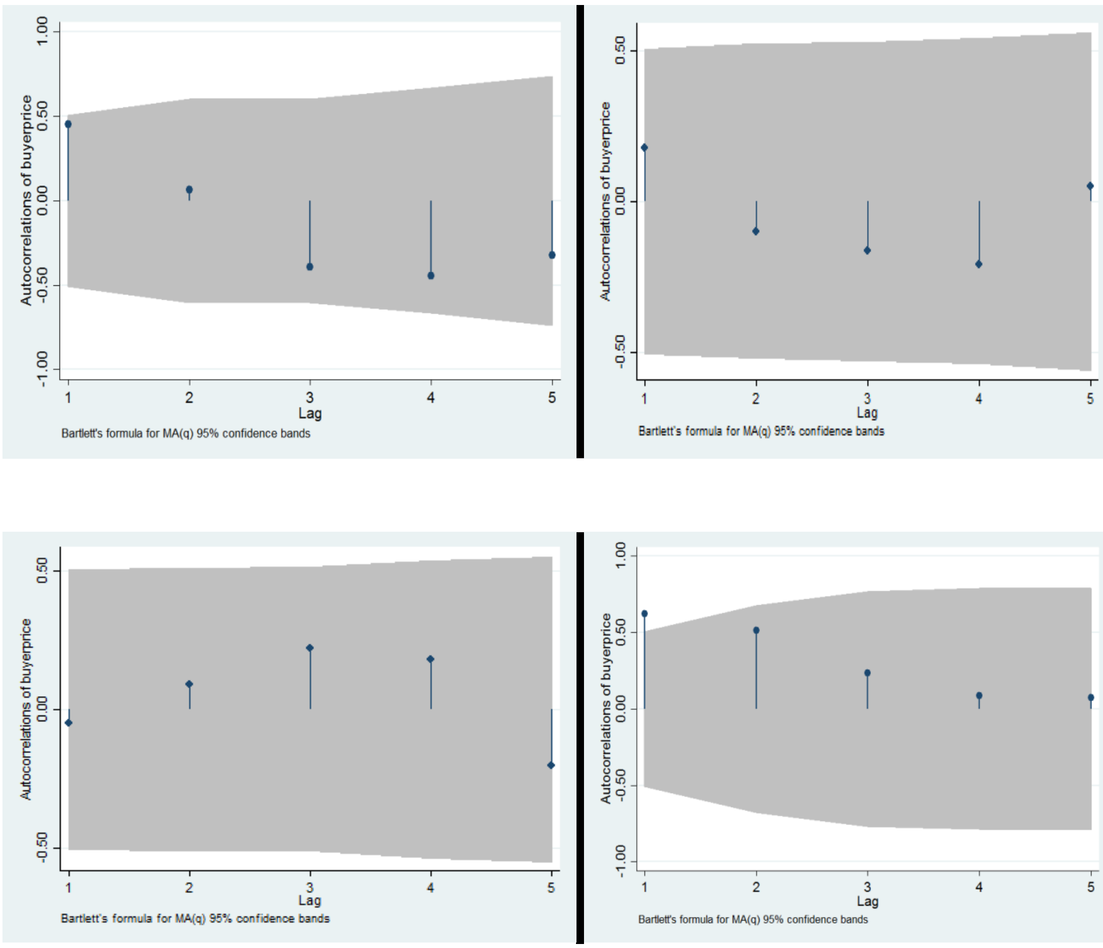


Figure A.2. Autocorrelation plots for the baseline treatments (last fifteen trading periods)

a) Double auction market with tax on buyer

b) Double auction market with tax on seller



c) Posted offer market with tax on buyer

d) Posted offer market with tax on seller