University of Arkansas, Fayetteville

ScholarWorks@UARK

Graduate Theses and Dissertations

12-2018

Three Essays on Corruption and Collusion

Arlo Redwine
University of Arkansas, Fayetteville

Follow this and additional works at: https://scholarworks.uark.edu/etd

Part of the Behavioral Economics Commons, International Economics Commons, and the Regional Economics Commons

Citation

Redwine, A. (2018). Three Essays on Corruption and Collusion. *Graduate Theses and Dissertations* Retrieved from https://scholarworks.uark.edu/etd/2971

This Dissertation is brought to you for free and open access by ScholarWorks@UARK. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of ScholarWorks@UARK. For more information, please contact scholar@uark.edu, uarepos@uark.edu.

Three Essays on Corruption and Collusion

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Economics

by

Arlo Redwine
University of Arkansas
Bachelor of Arts in English, 1999
University of Arkansas
Master of Arts in Economics, 2014

December 2018
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

Peter McGee, Ph.D.
Dissertation Director

Arya Gaduh, Ph.D.
Committee Member

Li Hao, Ph.D.
Committee Member

Abstract

This dissertation studies corruption and collusion with data derived from a laboratory experiment and household data. In Chapter 1 I study experimental procurement auctions with bribery and a public reserve to test for the tacitly collusive equilibrium described by Compte et al. (2005). Three sellers compete for 40 periods to sell a single item to a computerized buyer who accepts bribes and determines ties in bids and bribes randomly. In the closing periods, only 13.5% of auctions display the collusive equilibrium, but 58.7% of selling prices are noncompetitive. In comparison with simulated predictions for auctions that are corrupt but competitive, the mean selling price is 6.2% higher, efficiency is 35.2% lower, and the mean subject profit is 464% higher. Confusion leads to imperfect collusion, though some subjects learn to bid higher by observing bids. Men are more likely to bid the reserve. In Chapter 2 I present a method to detect corruption using only household data. I apply stochastic frontier (SF) analysis to measure the degree to which corrupt Chinese households underreport their income in comparison with other households, assuming the resultant differential is illegal income. Corrupt households on average underreport their income by 10%. I compare my results and method to those of Zhong (2018), who uses the same data but another method. Our results are similar, though only SF analysis 1) provides evidence of statistical significance, and 2) addresses endogeneity. My method provides an easy way to quantify the relative corruption between groups, regions, and countries. In Chapter 3 I apply the method of Chapter 2 to an Indonesian dataset. I find that the true incomes of public-sector households are, on average, about 50% higher than their reported income. I then divide the sample to support the findings of Martinez-Bravo et al. (2017), who exploit the fact that district mayors of the Suharto regime could finish their terms during the democratic transition, leading to exogenous variation in corruption exposure. When I restrict my sample to

shorter-exposed districts, my measurement falls to 37.1%; when I restrict my sample to longer-exposed districts, my measurement rises to 56.2%.

Table of Contents

1 Introduction

Testing theories in corruption and collusion with data can be difficult: the people who generate such data usually hide their actions. The three essays of my dissertation study these two topics with data derived from a laboratory experiment and household data. In both cases, the people who generate the data are unaware of my purpose, which produces an objectivity especially suited for testing theories.

In Chapter 1 I use a laboratory experiment to test the theory of Compte et al. (2005), who show that bribery in procurement auctions may lead to tacit collusion. This contradicts previous papers that describe conditions under which bribery and bidding lead only to a transfer from the government to the agent. These papers assume perfect competition in bribes. Compte et al., on the other hand, invoke an agent who has upper-limit reserves for both bribes and bids, determining ties in both randomly. The winner of the bribery contest pays a bribe, matches the lowest bid, and sells the item. A collusive equilibrium exists in which all sellers submit the maximum bribe and then allow the tie-breaking rule to determine the winner. I implement a version of Compte et al.'s model in the lab. Three continually rematched sellers compete for 40 periods to sell a single item to a computerized buyer who accepts bribes and determines ties in bids and bribes randomly. In the closing periods, only 13.5% of auctions display the perfectly collusive equilibrium, but 58.7% of selling prices are above the competitive price. In comparison with simulated predictions for auctions that are corrupt but noncollusive, the mean selling price is 6.2% higher, efficiency is 35.2% lower, and the mean subject profit is 464% higher.

In addition to testing the theory, I contribute to the more general literature on collusion in experimental auctions. Researchers in the lab have found scant evidence of tacit collusion among subjects who are continually rematched. Here limited bribery presents a coordinating device. If only a portion of bidders in this simplified environment try to take full advantage of this device, perhaps sellers in the real world, due to their bounded rationality, also fail to exploit collusive opportunities. Much of my paper explains why many subjects, having chosen to bid noncompetitively, attempt only imperfect collusion. First I exploit the results of a pre-experiment comprehension test to create a measurement of confusion. I show that unconfused subjects are about twice as likely to bid the reserve. Next, research has found that women tend to bid more aggressively and are more empathetic (Chen et al. 2013 and Toussaint and Webb 2005, respectively). I show that men are twice as likely to bid the reserve. Women are just as likely to be competitive, so the difference arises from a greater proportion of women choosing imperfectly collusive bids. Finally, subjects each period observe the bids of their two competitors, allowing them to learn from others. I find that among subjects who do not bid the reserve, those who observe one bid (two bids) higher than their own are more likely to bid higher in the next period compared with those who observe no higher bids (one higher bid).

In Chapter 2 I present a novel method to detect corruption using only household data. I apply stochastic frontier (SF) analysis to measure the degree to which Chinese households with opportunities for corruption underreport their income in comparison with other households, assuming the resultant differential is illegal income. I compare my results and method to those of Zhong (2018), who uses the same data but another method to find that the true incomes of households with corruption opportunities are, on average, about 15% higher than their reported incomes, an estimate that Zhong cannot statistically test. SF analysis produces an estimate of about 10% that is statistically significant. Given similarities in our robustness checks, my results vindicate Zhong's approach, though only SF analysis 1) provides evidence of statistical significance, and 2) addresses endogeneity, which may explain the estimate difference. My method provides a cheap and easy way to quantify the relative corruption between groups, regions, and countries. Most existing micro-level empirical analyses of corruption rely on administrative records, special-purpose surveys, or field experiments, which can be difficult or very costly to obtain. Most corruption studies in general rely on perception-based country-level corruption indices, which use corruption ratings based on expert opinions or surveys of business executives. Several studies, however, have found that perceptions of corruption are not perfect measures of actual corruption (e.g., Olken 2009 and Donchev and Ujhelyi 2014). Chapter 2 adds to this literature by exploiting a corruption-related question in the same household survey. Using both Zhong's method and the SF method to measure corruption, I find that households that report being less concerned about corruption live in areas with much more corruption. This may be the case because the least informed citizens live in the most corrupt areas — in which officials and the press do little to expose that corruption to the community.

In Chapter 3 I apply the method presented in Chapter 2 to an Indonesian dataset. I find that the true incomes of public-sector households are, on average, about 50% higher than their reported income, providing a rare measure of corruption's magnitude. I then divide the sample to support the findings of Martinez-Bravo et al. (2017), who exploit the fact that district mayors of the Suharto regime could finish their terms during the democratic transition. Suharto appointed these mayors in 1994, 1995, 1996, or 1997, leading to exogenous variation in corruption exposure. When I restrict my sample to shorter-exposed districts (those with mayors whose appointments were in 1996 or 1997), my corruption measurement falls to 37.1%; when I restrict my sample to longer-exposed districts, my measurement rises to 56.2%. My results provide a correlation absent from Martinez-Bravo et al. (2017): that between mayor exposure and corruption magnitude.

2 Chapter 1

Bribery and Tacit Collusion in Experimental Procurement Auctions

Abstract

In the lab I study procurement auctions with bribery and a public reserve to test for the tacitly collusive equilibrium described by Compte et al. (2005). Three sellers compete for 40 periods to sell a single item to a computerized buyer who accepts bribes and determines ties in bids and bribes randomly. In the closing periods, only 13.5% of auctions display the collusive equilibrium, but 58.7% of selling prices are noncompetitive. In comparison with simulated predictions for auctions that are corrupt but competitive, the mean selling price is 6.2% higher, efficiency is 35.2% lower, and the mean subject profit is 464% higher. Confusion leads to imperfect collusion, though some subjects learn to bid higher by observing bids. Men are more likely to bid the reserve.

2.1 Introduction

Government agencies often try to deter *explicit* collusion at their procurement auctions by implementing sealed-bid auctions (Carpineti et al. 2006), but what if the procurement agent accepts bribes?[1] Compte et al. (2005) show that such corruption (in the form of bribery) may lead to *tacit* collusion. This contrasts sharply with Beck and Maher (1986) and Lien (1986), who describe conditions under which bribery and bidding in thin markets lead only to a transfer from the government to the agent.[2] These earlier papers, however, fail to appreciate that competition in bribes is likely imperfect. Compte et al. model this imperfection through agents who are unwilling to accept bribes beyond a certain amount because higher bribes imply a greater probability of detection. Under this assumption, the authors show that corruption can lead to collusion at the government's upper-limit reserve price.

The importance of understanding the links between corruption and collusion is patent. Public procurement comprises about 15% of worldwide GDP (OECD 2008). Of the 427 foreign bribery cases concluded worldwide from mid-February 1999 to June 2014, 57% involved public procurement (OECD 2014). According to Compte et al., the primary policy implication of their findings is that the antitrust authorities should coordinate with criminal law enforcement. Each set of officials should be well-versed in the other's specialty. Compte et al. also show how the introduction of just one efficient honest bidder to a corrupt procurement auction can help restore competition. The policy implication is that governments should promote, or even subsidize, the entry of efficient firms.

[1] Sealed-bid auctions prevent sellers from observing deviations from collusive agreements, making it more difficult for them to sustain collusion.
[2] Beck and Maher even claim that under these conditions "controversies about the relative efficiency of bribery versus bidding may be moot."

Other recent papers also show how corruption in procurement can lead to efficiency losses. Burguet and Che (2004), for example, present a model in which sellers submit offers specifying quality and price. Simultaneously, they bribe the agent, who then manipulates the quality assessment to favor the high briber. If the agent's manipulation power is large, corruption is a device for facilitating collusion. Arozamena and Weinschelbaum (2009) show that in a single-item, first-price auction with no reserve price, corruption can change the competitiveness of uncorrupt bidders. An agent reveals to a favored bidder all rival bids and then allows the bidder to modify her original bid upward or downward.

In addition to a fuller understanding of the theoretical links between corruption and collusion, empirical evidence is needed. Compte et al. — as well as a related paper, Lambert-Mogiliansky (2011) — find corroboration for their theory in French authority reports and court cases.[3] But compared with rich countries like France, poor countries experience both corruption and collusion more intensely because of their less developed enforcement and judicial systems — the very institutions tasked with producing evidence.

I contribute to the empirical literature by implementing a version of Compte et al.'s model in the lab. For 40 periods I study auctions involving three continually rematched sellers and a computerized buyer who accepts bribes. Whereas Compte et al. make inferences based on indirect observations, the lab allows me to observe corruption directly. Only 13.5% of auctions in later periods display the perfectly collusive equilibrium of all sellers bidding the reserve, even though half of subjects attempt such collusion by bidding the reserve. Nevertheless, 58.7% of

[3] For example, Lambert-Mogiliansky (2011) cites the testimony of J.C. Mery, a Paris City Hall official. When he died, Mery left behind a videotape on which he describes how he accepted bribes from 1985 to 1994 in exchange for organizing collusion in the allocation of construction and maintenance contracts. The contracts generated up to 30% profit in an industry that averages 5%, despite the contracts' allocations to the lowest bidders.

selling prices are noncompetitive. Furthermore, compared with simulated predictions for competitive corrupt auctions, the mean selling price is 6.2% higher, efficiency is 35.2% lower, and mean subject profit is 464% higher. These changes, however, fall well short of those that would have prevailed had all subjects colluded at the reserve in every period. Confusion seems to explain much of this failure, though I find evidence that subjects learn from each other to raise their bids. In addition, men are nearly twice as likely to attempt collusion at the reserve.

This paper also contributes to the more general literature on price-matching schemes and collusion in experimental auctions. Researchers in the lab have found scant evidence of tacit collusion among subjects who are continually rematched. Here limited bribery presents a coordinating device. If only half of sellers in this simplified environment try to take full advantage of this device, perhaps sellers in the real world, due to their bounded rationality, also fail to exploit collusive opportunities. Yet, also like experimental subjects, some criminals may learn from each other to perfect their art. To hamper such learning policymakers should familiarize themselves with the mechanics of both bribery and collusion. This paper reveals and delineates those mechanics with a starkness possible only in the lab.

Büchner et al. (2008) is the only other lab experiment to study corruption and public procurement. Unlike Compte et al., the authors do not consider ex post collusion opportunities. Instead they assume that potential suppliers compete by simultaneously posting prices and offering bribes. The authors derive the optimal price-and-bribe bid, for which their experimental data provide qualitative support.

2.2 Theory and Experimental Design

Compte et al. consider a corrupt procurement agent who conducts a first-price sealed-bid auction for a single contract with a public reserve price $\overline{p}$ chosen by the government. There are n sellers indexed by i that bid for the contract. Seller costs c_i are drawn from distributions with positive and continuous density $f_i(\cdot)$ on $[\underline{c}_i, \overline{c}_i]$, known to the firms only. For convenience, order sellers so that $\overline{c}_1 \leq \overline{c}_2 \leq \cdots \leq \overline{c}_n$ and let $\underline{c} = \min_i \underline{c}_i$. Absent corruption, the agent simply allocates the contract to the seller with the lowest bid. In the case of ties, the agent chooses among lowest-bidding sellers with equal probability. Assuming that $\overline{c}_1 < \overline{p}$, in *any* Bayesian equilibrium the first-price auction allocates the contract at a price equal to at most $\min\{\overline{c}_2, \overline{p}\}$.[4]

With corruption, a second stage follows the bidding. The agent, after disclosing to all sellers the lowest bid, allows one seller to match the lowest bid and win the auction. The sellers compete for this favor by offering bribes, but only the winner actually pays a bribe. This competition is imperfect because the agent will not accept bribes above some threshold $\overline{B}$. If multiple sellers submit the highest bribe, the agent chooses among them with equal probability — unless one seller bids strictly below the others, in which case the agent chooses this seller with probability $1/n$ and the other(s) with equal probability. In this way, sellers cannot increase their chances of winning by bidding so low that some sellers cannot match the bid without the risk of losing money.

Compte et al. show that corruption facilitates collusion and thereby generates a price increase beyond the bribe amount. A Pareto dominant equilibrium exists in which all sellers bid $\overline{p}$ and offer $\overline{B}$ as a bribe. The agent then randomly picks one of the sellers as the winner. Compte et al.'s main result is as follows. Assume $\overline{p}$ is high compared to cost levels.[5] Then there exists a perfect Bayesian equilibrium in which the contract is sold at $\overline{p}$.

[4] Note that even if the densities f_i are important in deriving equilibrium behavior, Compte et al.'s results depend only on the bounds $\underline{c}$ and $\overline{c}_i$.

[5] Formally, assume $\overline{p} - \overline{c}_1 - \overline{B} > 0$ and $\overline{c}_2 - \underline{c} < (1/n)[\overline{p} - \underline{c} - \overline{B}]$.

Compte et al. present the following intuition for their result. Because competition in bribes stops at $\overline{B}$, if sellers compete only in bribes (and not in bids), they all profit in expectation as long as their cost parameters do not exceed $\overline{p} - \overline{B}$. Because the value of the bribe is bounded, any increase in bids translates into higher joint profits for the sellers. They thus have a joint interest in bidding as high as possible. Still, because sellers do not get the contract with certainty, some might compete in bids to increase their chances. But there is a high cost to doing so. For small bid deviations, bribe competition leads to ties because many sellers can propose $\overline{B}$ and still profit, and the agent need not pick the deviator with a larger probability. Thus, increasing the probability of winning would require decreasing the bid to a level at which other sellers cannot match the price without the risk of losing money. This level may be so low that each seller prefers to stick to the collusive outcome.

Compte et al. show that their findings are robust to more general models of imperfect bribery competition, as well as more complicated corruption mechanisms.[6] They also note that the assumption of a public reserve can be relaxed if one assumes that the corrupt agent reveals the reserve during the bribing stage (and presumably all firms bid above any possible reserve in stage one).[7]

Table 1 provides a summary of the experiment, which consisted of six sessions, each with nine subjects, conducted at the Behavioral Business Research Lab at the University of Arkansas. Subjects participated in three-person computerized auctions in which they were randomly

[6] Details can be found in unnumbered subsections of the paper's third section.
[7] Kagel (1995) note that "reserve prices, when they exist in practice, are typically not announced," adding that "when a reservation price is announced, it serves as a focal point for the collusive outcome."

rematched for 40 periods.[8] All subjects were sellers whose costs to produce a homogenous item, measured in experimental dollars to two decimal places, were drawn from $U[375.00, 400.00]$. Subjects knew that their competitors' costs came from the same distribution. All sessions used the same sequence of cost draws and group draws. I read the instructions aloud to ensure common knowledge. Subjects answered 18 computerized comprehension questions that provided explanations after each answer entry. Subjects also participated in three trial periods before the start of the paid periods. Earnings were given in experimental dollars (E$) with an exchange rate of E$40.00 = U.S.$1.00 for sessions lasting for 70 to 90 minutes (average earnings were U.S.$13.76, including a U.S.$5 show-up fee). Unless otherwise stated, all monetary figures are in experimental dollars.

Subjects first learned their cost privately and submitted an "initial offer price" that could be no higher than the reserve price of 475. They then learned the initial offers of their two competitors and entered a "resubmission fee" of up to 20.[9] The subject who offered the highest resubmission fee automatically matched the lowest initial offer and won the auction for this amount. The computer determined ties in fees randomly, and only the auction's winner actually paid the fee, which subjects knew. At the end of each auction subjects learned their "final offer price," which was potentially different from their initial offer only when they won the auction. If all subjects submitted zero as their fee, the subject with the lowest initial offer won, with ties determined randomly.[10]

[8] I chose random rematching in part to match the static theory of Compte et al., who do not describe how sellers reach the collusive equilibrium. Also, in comparison with repeated play, random rematching makes collusion in the absence of communication (or limited bribery) much less likely (Kagel 1995).
[9] Subjects could choose values up to two decimal places for both the initial offer price and the resubmission fee.
[10] All subjects submitted zero as their fee in only 5 of 720 auctions (0.7%).

The cost and reserve parameters meet Compte et al.'s equilibrium conditions. They also correspond to a highest possible profit margin of 25.3% and a maximum resubmission fee of 4.2% of the reserve. According to Compte et al., the French Ministry of Equipment considers rejecting all bids when the lowest bids exceed its own cost estimate by about 25%. Recall that Compte et al.'s general theory describes an agent who punishes sellers who knock out other sellers. Given my specific cost distributions and parameters, however, this rule is not necessary to sustain the equilibrium of all sellers bidding 475 and submitting 20.[11] Subjects cannot profitably deviate even if they draw a cost of 375 while believing both competitors draw the highest realization of costs, 400.[12] Under these circumstances subjects could bid 419.99 and earn 24.99 with certainty, but they could earn 26.67 in expectation by not deviating. Of course there is no reason for subjects to believe that their competitors draw particular costs. If subjects bid according to the risk-neutral Nash equilibrium described in the next section, subjects who draw 375 earn only 8.33 when they win. Furthermore, many subjects are probably risk-averse, leading to an even lower competitive profit and less incentive to deviate from the collusive equilibrium.

I chose to reveal to subjects their competitors' initial offer prices both before and after they submitted their fees. Thus subjects could use their bids as signals of collusive intent. The corrupt agent of the model might encourage such signaling, given that collusive sellers are more likely to afford maximum bribes. A history table at the bottom of the computer screens included the subject's past costs, initial offers, final offers, and profits, along with the competitors' past initial offer prices (all in experimental dollars). The experiment was programmed in z-Tree (Fischbacher 2007). See the Appendix for the instructions and screenshots of the user interface.

[11] I also chose symmetric sellers and a uniform distribution to allow for standard predictions and simpler instructions, though research has shown that symmetry fosters collusion.
[12] Like Compte et al., I assume risk-neutrality and that subjects submit the highest fee they can afford.

2.3 Predictions and Results

2.3.1 Nomenclature

Table 2 summarizes the nomenclature of the analysis. To make comparisons with the experimental outcomes, I often simulate the outcomes of "honest auctions," i.e., auctions with the same (relevant) parameters as the corrupt auctions but with no bribery stage. Also for comparison's sake, I often simulate the outcomes of "competitive auctions": corrupt auctions in which, as I explain later, all subjects add 20 to the bids they would have submitted in an honest auction.[13] Note that subjects in honest auctions also act competitively, but I use the term only when referring to subjects in corrupt auctions who augment their bid in the manner described. "Perfectly collusive auctions" are corrupt auctions in which all three subjects bid 475, leading to a selling price of 475. "Imperfectly collusive" auctions are corrupt auctions in which the selling price is above 420 but below 475. For reasons I explain later, I classify bids and selling prices into three groups: "competitive" if they are 420 or below, "imperfectly collusive" if they are above 420 but below 475, and "perfectly collusive" if they are equal to 475.

2.3.2 Predictions

For honest auctions the risk-neutral Nash-equilibrium (RNNE) bidding function is $bid = (2/3) \times cost + 400/3$, the expected selling price is 387.5, and the expected profit is 2.08. In previous experimental auctions (Kagel and Levin 2014) subjects bid slightly lower (or higher in the case of forward auctions). Efficiency, defined as the percentage of auctions won by the lowest-cost subject, should be 1.00.

[13] I simulate the outcomes of honest auctions by transforming the cost draws of the experiment using the Nash equilibrium bidding function for risk-neutral sellers (defined in the next subsection). I simulate the outcomes of competitive auctions by transforming the cost draws with the same bidding function shifted upward by 20, and by assuming subjects submit the highest resubmission fee they can afford.

With bribery, competitive subjects may react mechanically by increasing their bid by the maximum resubmission fee of 20. In comparison with honest auctions, both the RNNE bidding function and predicted selling price shift upward by 20. That is, the function becomes $bid = (2/3) \times cost + 400/3 + 20$, the same function for an honest auction in which costs are drawn from $U[395.00, 420.00]$. The expected selling price rises 5.2% to 407.5. The predicted net-of-fee selling price, 387.5, does not change. The resubmission-fee mechanism simply transfers 20 from the government to the agent. The predicted efficiency, however, is no longer 1.00. To see why, consider a set of three cost draws in order of increasing magnitude: C_1, C_2, and C_3. If subjects with these draws compete in a corrupt auction using the bidding function described, the lowest-cost subject wins with certainty only when his bid is so low that neither competitor can submit a fee of 20, match the minimum bid, and not lose money. Otherwise, each subject who can afford a fee of 20 has an equal chance of winning. The condition under which the lowest-cost subject knocks out both competitors is $3C_2 - 2C_1 > 400$.[14] A similar expression describes the condition under which the subject knocks out only one competitor. A simulation of 300,000 auctions indicates that the probability of knocking out one competitor is 0.45, and the probability of failing to knock out either competitor is 0.11.[15] Given these probabilities, the expected efficiency is 0.70. The same simulation shows that the mean subject profit is 1.78, down from the 2.08 that would have prevailed if it weren't for the efficiency loss. In this regard, sellers actually earn more with honest auctions than with corrupt auctions that are competitive. Again, when I refer to competitive auctions, I mean those in which all three sellers add 20 to the RNNE bids

[14] The lowest-cost subject hopes to bid low enough to knock out the subject with the second lowest cost (naturally knocking out the other competitor as well). This happens when $(2/3) \times (C_1 + 20) + 140 - 20 - C_2 < 0$. Rearrangement of terms leads to $3C_2 - 2C_1 > 400$.

[15] I created 900,000 observations, divided into 300,000 groups, along with a cost draw and bid for each observation. Then I found the lowest bid of each group (*minbid*), and for each group counted the number of 20-bribers, i.e., observations in which $minbid - cost - 20 \geq 0$. I chose a winner randomly among these 20-bribers.

they would have offered in an honest auction. To better organize my results, however, I broaden my definition of competitive bids to 420 or below because the highest possible bid under this strategy is 420.

Competitive subjects choose the highest resubmission fee they can afford, and they have a chance of winning only if they can afford 20. In the simulation, the mean fee offer is 17.22 with a standard deviation of 4.13 and a minimum of 3.34. The percentage of fee offers equal to 20 is 55.6%.

Subjects also may bid noncompetitively in the hope that their two competitors do likewise. Once subjects commit to collusive bids (above 420) the profit-maximizing bid is 475, which I call perfectly collusive. Bids that are above 420 but below 475, which I call imperfectly collusive, do not increase a subject's chances of winning and can potentially lower the prize for which the subjects are competing. If all three subjects of an auction bid 475, then the expected selling price, efficiency, and subject profit are 475, 0.33, and 22.50, respectively, and the expected fee offer is 20 — as it is when all subjects bid 420 or above.

When all three subjects bid either competitively (according to the function) or collusively, they have an equal chance of winning. On the other hand, some subjects may bid competitively while others bid collusively. According to simulations, when only two subjects bid competitively, their chances of winning increase to 0.38. A subject who alone bids competitively wins with probability 0.50. Behaviorally, the most interesting questions in the lab are whether subjects choose to bid competitively or collusively, and if the latter, whether they realize that the most sensible bid is 475 — a realization that requires only one step of backward induction. Other questions are whether subjects choose the highest fee they can afford, and whether some subjects learn to bid higher by observing competitors' bids. In sum, I test four hypotheses:

Hypothesis 1: Subjects submit the highest resubmission fee they can afford.

Hypothesis 2: When subjects bid 420 or below, they bid according to $bid = (2/3) \times cost + 400/3 + 20$.

Hypothesis 3: When subjects bid above 420, they bid 475.

Hypothesis 4: Among subjects who do not bid 475, those who observe one bid (two bids) higher than their own are more likely to bid higher in the next period compared with those who observe no higher bids (one higher bid).

2.3.3 Main Results

I wish to study how behavior changes during the sessions, so I focus on the opening periods and the closing periods. Specifically, I look at periods 1-7 and 34-40, with the data from the latter acting as a proxy for an equilibrium.[16, 17] For these period ranges, Table 3 reports means for the bid, the selling price, subject profit, and efficiency, and makes comparisons between these values and the predictions for competitive auctions and perfectly collusive auctions. These predictions are the result of simulations using the same sequences of cost draws. As a robustness check, Tables A1 and A2 in the Appendix reproduce Table 3 using, instead of the first and last seven periods, the first and last five periods, as well the first and last 10 periods.

Because Compte et al. present a static theory, statistics for the final periods are my main results, and I refrain from reporting these figures for all periods. Thus consider only the 126

[16] Because I rematched subjects from period to period, only observations at the session level are independent. Standard deviations thus capture variation among the six sessions. In what follows, I report all *p*-values using Wilcoxon signed-rank tests. Almost all main results are also significant according to more conservative sign tests.
[17] To decide which opening and closing periods to study, I examine the overall mean bid (using the six session observations of the mean) for blocks of periods at both ends of sessions. The mean for the first 20 periods is not statistically different from the mean for the last 20 periods. I continue to reduce the periods from both ends until I discover a difference at the 5% level.

auctions of periods 34-40. Compared with predictions for honest auctions, the mean selling price is 11.7% higher, efficiency is 54.0% lower, and mean subject profit is 359% higher. All differences are significant at the 5% level. Compared with predictions for competitive auctions, these same differences are as follows: 6.2% higher, 35.2% lower, and 464% higher, respectively, all significant at the 5% level. Clearly collusive bidding takes place. But the magnitudes of these percentage changes fall well short of those that would have prevailed had all subjects colluded perfectly (bid 475) in every period, in which case (using further simulations) the same percentage changes would have been 16.5% higher, 60.6% lower, and 1,240% higher when compared with the competitive predictions.

Nevertheless, for periods 34-40, the modal bid is 475, which is the bid of 186 of the 378 decisions (49.2%, Figures 1-2 and Table 4). The second most common bid is 450 (4.8% of decisions), followed by 440 (4.0%) and 420 (3.7%). On the other hand, bids of 420 or below are prevalent in the closing periods: 16.4% of bids. The remaining bids, 34.4% of the total, are above 420 but below 475. As I noted earlier, such choices are imperfectly collusive in the sense that they potentially limit profits for no obvious reason. So perhaps confusion is a bigger hindrance to perfect collusion than competitiveness. It bears emphasizing, however, that 83.6% of bids are noncompetitive.

Given the heterogeneity in bidding behavior, only 17 of the 126 auctions (13.5%) display the perfectly collusive outcome in which all subjects bid 475 and submit 20.[18] But for period 40 alone this percentage doubles to 27.8% (5 out of 18 auctions), and this behavior still leads to a modal selling price of 475, which is the cost of procurement for 18 of the 126 auctions (14.3%,

[18] In only one auction of the closing periods did all three subjects submit 20 and bid the same initial offer not equal to 475: an auction in which all subjects bid 445. In only one auction did all three subjects submit the same fee not equal to 20: an auction in which all subjects bid zero due to one subject's bid of only 2.00 above cost.

Figures 1-2 and Table 4).[19] The second most common selling price is 440 (7.1% of auctions), followed by 420 (5.6%) and 430 (4.8%). Competitive bidding has a strong effect: 40.5% of selling prices are 420 or below. This leaves 45.2% of selling prices in the region between 420 and 475. In sum, 58.7% of selling prices are noncompetitive, and 475 is still modal. Figure 3 shows the kernel densities of bids and selling prices, illustrating how bids skew leftward while selling prices skew rightward.

The difference between collusive attempts and outcomes is further revealed when examining the changes in overall means from the opening periods to the closing periods. The mean bid grows significantly ($p = 0.046$), as shown in Figure 4; the means for the selling price and profit do not ($p = 0.116$ and $p = 0.176$, respectively). In the last subsection I investigate whether learning can partly explain why the mean bid changes.

2.3.4 Resubmission Fee Behavior

Finding 1: Subjects usually submit the highest resubmission fee they can afford when that fee is 20.

Hypothesis 1 is that subjects submit the highest resubmission fee they can afford. The data for periods 34-40 include 378 fee choices, 286 of which (75.7%) are 20. All subjects who bid the maximum can afford to do so; however, an additional 5.3% of fee decisions could have been 20 without risk of losing money.[20] The bids tied to these latter fee choices are more likely to be imperfectly collusive: 20.0%, 60.0%, and 20.0% of these bids are competitive, imperfectly collusive, and perfectly collusive, respectively, compared with 16.4%, 34.4%, and 49.2% for all

[19] The last-period jump may be an end-game effect resulting from last-ditch efforts to win big. However, the percentage of bids equal to 475 for periods 34-40 and period 40 are similar: 49.2% and 53.7%, respectively. So the jump may be more the result of chance.

[20] Two of these subjects won anyway, raising their profit on average by 7.50 relative to the profit they would have made had they chosen 20; the others, by not bidding 20, forwent their chance of winning an average profit of 27.65.

bids of the final periods. In regard to the remaining 72 fee choices (19.0% of the total), in which subjects can afford at most a value below 20, every fee choice is lower than the highest affordable fee except for seven instances in which both the most affordable fee and the actual fee are zero. In 68.1% of these remaining choices, subjects cannot afford the maximum because of a competitor's bid. Otherwise, the subject's own low bid precludes a fee of 20. Thus, unsurprisingly, the bids corresponding to these 72 fee choices are unusually competitive: 40.3%, 30.5%, and 29.2% are competitive, imperfectly collusive, and perfectly collusive. For all 40 periods Figure 5 compares the mean fee offer with the mean highest fee offer subjects can afford. Table 5 shows that compared with these two means for periods 1-7, these means for periods 34-40 are much closer in value while still remaining distinguishable. Recall that subjects choose their fee after learning the lowest bid, i.e., the selling price. These subjects, therefore, may be choosing a fee corresponding to some minimum acceptable profit (instead of zero) in the off-chance they win.[21] Also, subjects often heuristically choose a multiple of five, perhaps in lieu of bothering to calculate the highest affordable fee.[22]

Table 5 also shows that the mean winning fee is slightly lower than predicted. I discuss later how fee behavior relates to bid behavior. In particular, subjects theoretically should choose a fee below 20 only after competitive bidding, to which I now turn.

2.3.5 Competitive Behavior

Finding 2: When subjects bid 420 or below, they do not bid according to $bid = (2/3) \times cost + 400/3 + 20$, but the function's slope is qualitatively predictive.

[21] In 16 of the closing 126 auctions (12.7%), the winner does not submit a fee of 20. The mean fee of these auctions is 8.85 with a range of 0 to 17.5. The mean selling price is 395.00 with a range of 376.5 to 405. Recall that according to both the competitive and collusive theories, the winning fee should always be 20.

[22] For this heuristic Lynn et al. (2013) reviews both the evidence and the explanations.

Because I later explore how subjects learn during sessions, I often expand my analysis to all periods in addition to just periods 34-40. Here I consider both ranges. For all periods 18.0% of bids are 420 or below, a percentage only slightly higher than the 16.4% reported for the closing periods. If subjects had adhered to the competitive bidding function, because of the uniform cost distribution, the distribution of these bids would have approached uniformity between 403.33 and 420. Figure 6's histogram shows this is not the case (the histogram for periods 34-40 is similar). Subjects disproportionately choose bids that are multiples of five — in particular, 24.2% of bids in this region are 420. Furthermore, 24.0% of the bids are below the theoretical lower limit of 403.33. For all periods (alternatively, periods 34-40), the proportion of bids in the 420-and-below region that are lower than the bidding-function predictions is 42.5% (48.4%), but this proportion grows to 56.1% (62.5%) when I consider only bids strictly below 420. Figure 7 shows line graphs of the mean bid and mean predicted bid for values strictly below 420. The actual mean is always lower. The mean bid and mean predicted bid are uncorrelated across periods, and the bids summarized by the means are very weakly correlated, with a Spearman coefficient of 0.19 ($p = 0.001$).

If all bidding had been collusive, presumably there would have been no correlation between bids and costs. Yet the Spearman correlations for periods 34-40 of two sessions are statistically significant (though weak): 0.31 and 0.32 (p-values ≤ 0.014). Seven subjects (13.0%) among five sessions have significantly positive coefficients for the closing periods, and the correlations are strong: at least 0.73 (p-values ≤ 0.060). For all periods the correlation for 16 of the 54 subjects is 0.30 or above (p-values ≤ 0.053). As an example of a competitive subject, Figure 8 shows the bids of a subject in the fifth session whose correlation coefficient for all periods is 0.72 ($p = 0.000$) and mean bid 412.63. The competitive predictions capture the subject's choices well, especially in the closing periods. The bids for a significant percentage of other subjects follow a pattern similar to the one shown here, at least for some periods.

Bid-cost correlations do not indicate necessarily that subjects are bidding low to knock out competitors. For 7 of the 16 subjects above, correlations are greater than 0.30 and significant even for bids above 420. Recall that there is no reason for subjects to react to costs when bidding above 420. Despite this caveat, I continue to use the 16 subjects as a proxy for competitive bidders. Competitive bidding leads to very low profits in terms of real dollars. Asymptotically the expected winner's profit per auction is 6.25, which translates to $0.16. In comparison, sellers who win after bidding 475 can expect to earn 67.5, which translates to $1.69. As a group the 16 competitive subjects on average earn $10.62 during a session; the other 38 subjects earn $15.08. Subjects may persist in competitive behavior despite the low profits because they disregard the exchange rate, or they experience spite or the joy of winning.[23]

Compte et al. examine the effect of introducing to their model an efficient "outsider," a firm that cannot bribe.[24] As long as the upper limit of the outsider's cost support is below the upper limit of the most efficient corrupt seller, the selling price is below this corrupt seller's upper limit in any Bayesian equilibrium. The outsider is ready to harden price competition due to its exclusion from the bribery stage. By proposing a low enough price, the outsider can ensure that other less efficient firms cannot afford to compete in bribes. In a similar fashion, even though

---

[23] For example, a post-session survey asks subjects to describe their strategies for choosing bids and fees. One subject writes: "I played a price close to my cost to get people to not bid and I would sell the item. I took a small profit over 0 profit. And if my competitor won, they would get very little."

[24] The authors also show that increasing control over the agent, i.e., reducing the upper limit for bribes, does not reduce (and can even increase) the ability of firms to collude. Controls on efficient firms may partly restore competition in prices, assuming the agent cannot sufficiently favor a bid from a briber based on something other than price.

corruption shifts bids upward, the few subjects who persist in competitive bidding severely limit session-wide profits.

To study more precisely how subjects respond to costs, I regress bid on cost and session dummies. As other papers have done with similarly structured experimental data (Engelbrecht-Wiggans et al. 2007 and Fugger et al. 2015), I use a Tobit approach to account for censoring at both costs and the reserve, and I account for dependence among bids by using random effects. I estimate the following model:

$$bid_{it} = \alpha + \beta_1 cost_{it} + \sum_{j \neq 1} \beta_j session_j + \eta_i + \varepsilon_{it}.$$

Note that there are two error components — one that is independent across all observations, ε_{it}, and one that is subject-specific (the random effect), η_i. Both error terms have a mean of zero and a positive standard deviation. Recall that for subjects who bid competitively, the predicted marginal effect is 0.67, but subjects in the literature tend to bid more aggressively. The five columns of Table 7 report marginal effects for all bids, bids of periods 1-7, bids of periods 34-40, bids above 420, and competitive bids. All but one effect is significant at the 1% level (p = 0.012 for bids above 420), and differences in the effect corroborate the correlation findings. The effect declines to 0.25 from 0.43 from the opening to the closing periods, with an effect of 0.34 for all periods. The effect is 0.10 for bids above 420, and 0.44 for competitive bids. For bids strictly below 420 and after the removal of one extremely low bid (304.76), the effect grows to 0.53 (p = 0.000). In the next subsection I show that the effect grows further after the addition of covariates.

As mentioned earlier, competitive bidding affects fee choice. For all periods, in 58.4% of cases in which subjects do not submit 20, they cannot do so without the risk of losing money.

This percentage translates to 17.6% of all fee choices.[25] Subjects who bid competitively increase their chances of winning from 0.31 to 0.43 ($p = 0.035$), with the latter probability having a wide range among sessions: 0.33 to 0.52.[26] A post-session survey asks subjects to describe their strategies for choosing bids and fees. I cannot clearly classify many of the answers. Several subjects mention adding 20 to their preferred initial offer, or bidding "safely," but these subjects may be operating under alternate rules. Some subjects are unambiguously competitive. For example:

> I decided to use lower initial offers in order to beat out my competition. I found that I won more often when I came in with a low offer than when I tried to randomly be selected to win the battle with everyone having high initial offer prices.

Even more subjects clearly state the perfectly collusive strategy, which I address next. For example:

> Why put anything less than $475 and $20? You have a 1/3 chance of winning. People are too dumb to get that though.

2.3.6.a Collusive Behavior

Finding 3: When subjects bid above 420, they bid 475 about three-fifths of the time.

For periods 34-40, the value of 58.9% of bids above 420 is the reserve. The corresponding percentage for all periods is 48.5%. There is strong evidence that some subjects do not understand the auction rules, or at least doubt their understanding. Subjects, having once bid 475, often deviate in subsequent periods by offering lower noncompetitive amounts, starting with

[25] More specifically, in 12.5% of fee choices subjects cannot bid 20 safely due to a competitor's low bid (as opposed to their own).

[26] Recall from Section 3.2 that when only two subjects bid competitively, their chances of winning increase to 0.38. A subject who alone bids competitively wins with probability 0.50.

474.99. The average run length of consecutive 475 bids is 7.4 periods (s.d. = 2.05). More than half of runs (54.2%) are just one period. Another 14.4% are only two or three periods. Many subjects don't seem to understand that their bid, when it is beyond 420, has no bearing on their chances of winning.[27]

2.3.6.b Confusion

The post-session survey asks subjects to rate the statement "The instructions for the experiment were clear and easy to follow" on a five-point Likert scale from "strongly agree" to "strongly disagree." Nobody strongly disagrees, but nine subjects disagree, and eight subjects neither agree nor disagree. Reported confusion may be a poor proxy for actual confusion, so I also report the number of comprehension questions that subjects missed on their first try. On average, subjects missed 3.5 questions out of 18 (s.d. = 2.40) with a range from zero to nine. I construct a measurement of actual confusion: an indicator of whether subjects missed four or more comprehension questions, as 23 subjects (42.6%) did. I will refer to these subjects as "confused." Table 6, which takes all 40 periods into account, shows that unconfused subjects are about twice

[27] The experimental literature on deception reports instances "incomplete cheating" (see Hao and Houser 2017 for an overview). Although I never insinuated to subjects that collusive bidding was cheating, I cannot rule out the possibility that some subjects held this view and refrained from "complete cheating" (bidding the reserve).

as likely to bid 475, and about twice as *unlikely* to bid competitively (p = 0.075 and p = 0.028, respectively). [28, 29, 30, 31]

Partly driving results are subjects who alternate between collusive and competitive bids. Among subjects who bid above 420 and then lower their bid in the next period, 34.8% lower to 420 or below. This percentage drops to 18.5% among subjects who bid 475 in the previous period. Among subjects who bid 420 and below and then raise their bid, 66.7% raise to above 420. Occasionally subjects seem to decide their strategy based on their cost draw. Those who draw 387.5 and below are twice as likely to bid competitively, taking advantage of their low cost draw, as those who draw above 387.5: 0.24 vs. 0.12 (p = 0.028).

2.3.6.c Gender Effects

Subjects are comprised of 24 men and 30 women. Research has found that women tend to bid more aggressively (Ham and Kagel 2006, Chen et al. 2013). Here aggressiveness might matter not only for subjects who bid competitively, but for subjects who bid above 420 but below 475, because confused subjects (even if not measured as such) are presumably bidding in this region.

[28] A subject's missing of at least four test questions is above mean by at least one-fifth of a standard deviation. As a robustness check, I redefine my confusion measurement as the missing of at least five questions, representing three-fifths of a standard deviation. The number of subjects measured as confused drops 30.4% to just 16. The proportion of unconfused subjects who choose 475, 0.44, remains above the proportion of confused subjects who choose 475, 0.34, but the difference is not significant (p = 0.250).

[29] I recognize that my measurement of confusion could be interpreted as a measurement of other attributes, such as attentiveness or intelligence. It is also possible that quiz performance and competitiveness are directly correlated, though that story is less convincing. Of the 23 subjects measured as confused, 10 describe a competitive strategy in the post-session survey, 10 describe a collusive strategy, and three are ambiguous. Only four confused subjects state a clear misunderstanding of the rules, half of whom describe a collusive strategy and half of whom describe a competitive strategy. These observations, combined with the description of how confused and unconfused subjects actually behave, imply that some confused subjects learn the rules (or change preferences) during the course of the experiment. In the final periods unconfused subjects are still more likely to bid 475, but confusion no longer predicts competitiveness (see Footnote 30).

[30] For periods 34-40, unconfused (confused) subjects bid 475 with probability 0.59 (0.38, p = 0.028). Unconfused subjects bid competitively with probability 0.11, indistinguishable from confused subject's 0.25 (p = 0.206).

[31] In the Appendix are the results of one session in which subjects participate in honest auctions. Standard predictions capture the results well, providing evidence that subjects are capable of understanding the novelty of a procurement auction.

Another reason why women might be more inclined to bid in this region is that they feel greater pity for the computerized buyer — that is, the researcher whose budget declines more under perfect collusion than under imperfect collusion.[32] Table 6 reports that for all periods men are nearly twice as likely, 0.54 versus 0.29, to bid 475 (p = 0.046).[33] Furthermore, 30.0% of women never bid 475, while the corresponding figure for men is just 12.5% (p = 0.140). The end result of both confusion and gender effects is that unconfused males bid the reserve 69.9% of the time. Women are just as likely to be competitive, so the difference arises from a greater proportion of women choosing imperfectly collusive bids: 0.53 vs. 0.29 (p = 0.028). The proportions of women measured and self-reported as confused are greater than those for men — 0.49 versus 0.39 and 0.37 vs. 0.23, respectively — though these differences are not significant. Nor are the differences in the mean total profit and the mean selling price, partly because the winner's bid becomes the price in just 43.2% of auctions. Mean bids, both unconditionally and in the competitive and imperfectly collusive regions, also show no difference by gender. But a difference does arise (p = 0.046) for bids greater than 420: men on average bid 464.81 and women 456.20.[34] Women who choose imperfectly collusive bids tend to bid high values more often: 56.9% of their choices in the imperfectly collusive region are 450 or higher, compared with men's 47.1%. When considering all bids, women are much more likely to bid 450: 0.09 vs. 0.05. Finally, research has found men to be more confident than women in strategic situations (Croson and Gneezy 2009). There is evidence that this is occurring here: men and women who bid 475 do so again in the next period with probability 0.94 and 0.76, respectively (p = 0.028).

[32] Toussaint and Webb (2005), citing several papers, note that "empirical researchers have found that gender differences in empathy commonly indicate that women have higher levels than do men."
[33] For periods 34-40, men (women) bid 475 with probability 0.66 (0.36, p = 0.046).
[34] For periods 34-40, these same figures for men and women, respectively, are 467.49 and 460.06 (p = 0.075).

2.3.6.d Regression Analysis

I continue the Tobit regression analysis from section 3.3. I regress bid on cost, the ex ante confusion indicator, the period, session dummies, and gender.[35] The five columns of Table 7 report marginal effects. As expected, the average male bid is 8.75 higher. Men who bid competitively, however, on average bid 3.97 *lower*. Confused subjects on average bid 15.5 lower, and confusion is not significant for competitive bids. Bids on average grow 0.30 per period (or 11.70 during a session). Changes between the opening periods and the closing periods also are as expected: the period effect drops drastically in both magnitude and significance; the cost effect drops to 0.26 from 0.47; the gender effect is insignificant for the opening periods but is significant (p = 0.031) and strong (12.26) for the closing periods; and confusion's effect drops in both significance (p = 0.000 to p = 0.065) and magnitude (-18.79 to -10.33). For competitive bids, the marginal effect for cost is 0.45. For bids strictly below 420 and after the removal of the extremely low bid mentioned earlier, the effect grows to 0.55 (p = 0.000), with 0.67 as the upper bound of the 95% confidence interval.

2.3.6.e Fairness

Feelings of unfairness also may have influenced behavior. After all, in most auctions the price is not the winner's bid. Usually winners settle for a lower price set by a competitor, with an average lowering of 19.03. One subject at the end of the experiment voiced irritation, saying he didn't understand why others chose to decrease payouts by bidding below 475, in some cases for "pennies." The post-session survey also asks subjects to rate the statement "The experiment was fair" on a five-point Likert scale from "strongly agree" to "strongly disagree." Two subjects strongly disagree, four subjects disagree, and eight subjects neither agree nor disagree. In

[35] When I add gender/confusion interactions, none of accompanying coefficients are significant.

response to the strategy question, some subjects describe prices that are “too high,” but the subjects may be fairness-minded or confused. No subject mentions fairness per se.

2.3.7 Learning Effects

Finding 4: Evidence supports Hypothesis 4.

Some subjects seem to learn. One subject, for example, writes:

> At first I did not understand that I did not need to have the best offer in the first turn. After the practice, I decided to stay with 475 as it would not decrease the selling price.

Many subjects immediately understand the collusive strategy, submitting bids that act as signals. Figure 9 shows how 475-bidding (and 20-submitting) grew significantly during the sessions, contributing to the steadily growing mean bid of Figure 4. In periods 1-7, the mean bid is already 444.96 (s.d. = 12.29). Subjects bid 475 about a quarter of the time and 450 another 11.1% of the time. In periods 34-40, subjects bid 475 nearly half the time and 450 only 4.8% of the time.

In the opening seven periods, 22 out of 54 subjects (40.7%) choose the 475/20 pair at least once, and three choose it every period. In the last seven periods, 33 subjects (61.1%) choose the pair at least once, and 12 choose it every period ($p = 0.035$ for the comparison between percentages). For all periods, 40 subjects (74.1%) choose the pair at least once, with five subjects bidding 475 in every period.

Each period subjects receive two types of signals: their winning or losing, and the bids of their two competitors. In regard to winning and losing, the resubmission fee is paramount: subjects who submit 20 increase their probability of winning from 0.14 to 0.41 ($p = 0.028$).[36] Among subjects who did not submit 20 in the previous period, those who lost submit a higher fee

[36] Also recall that choosing a competitive bid increases the probability of winning by 0.12.

with probability 0.80, compared with 0.66 for those who won ($p = 0.046$). In regard to competitors' bids, Table 8 reports the probabilities that subjects raise or lower their bids after observing bids in the previous auction. My approach borrows from the learning direction theory devised in the late 1990s (Selten 2004) and further developed in recent years (see Crawford 2013 for an overview). In the table, a "higher" ("lower") signal is competitor's bid in the previous auction that was higher (lower) than the subject's own bid. Subjects who did not bid 475 in the previous period and faced two lower signals, bid higher with a probability of only 0.27. This probability rises to 0.41 if the subject faced one higher signal ($p = 0.075$) and to 0.59 if the subject faced two higher signals ($p = 0.028$, for the comparison between 0.41 and 0.59). Furthermore, subjects are equally likely — with a probability of about 0.40 — to bid lower after facing either two lower signals or one higher signal. But this probability drops to just 0.29 after facing two higher signals ($p = 0.028$). Table 9 re-creates Table 8 using only bids greater than 420. As expected, the learning effects are intensified: the probability of bidding higher after experiencing two higher signals is now 0.71. The last row of both tables reports that if subjects *did* bid 475 in the previous auction, the chances they will decrease their bid in the current period is roughly 0.10, regardless of signals.

Signals of winning and losing interact with the higher and lower signals only to a degree. Every probability for a movement toward 475 in Tables 8 and 9 is higher when interacted with winning, but these differences are not statistically significant. Losing subjects, however, are significantly more likely to lower their bids in a period after not bidding 475, regardless of signals. Unconditionally, losing subjects who did not bid 475 in the previous period lower their bids with probability 0.40; winning subjects with probability 0.24 ($p = 0.028$ for the comparison). Are losing subjects more likely to lower their bids even when their bids are above

420? In this case, lowering their bids to an amount still above 420 cannot help them win and could actually lower their profit. I find that subjects do respond to these false signals. For bids above 420 only, the probability that subjects lower their bids after losing is 0.27, compared with 0.19 for winners ($p = 0.075$).

During the course of most sessions the mean bid grows, and so does the number of bids equal to 475. Subjects who persist in submitting low bids, therefore, face a greater number of higher signals toward the end of a session. Because experience and number of signals are correlated, perhaps the changes in trajectory likelihoods reflect introspection instead of signaling. However, the probability that subjects who did not bid 475 in the previous auction bid higher does not change during the sessions when I examine various blocks of periods from both ends of sessions. In fact, a plot of mean probabilities by period shows a slight downward trend (Figure A1 in the Appendix). Furthermore, reproductions of Tables 8 and 9 using data from periods 1-20 and then from periods 21-40 result in only minor differences in likelihood estimates. I also estimate a linear probability model, regressing the likelihood of bidding higher on cost, period, and the two types of higher signals, assuming the subject did not bid 475 in the previous auction. I used, at the individual level, fixed effects and clustered standard errors. The period coefficient is highly insignificant ($p = 0.527$), while the coefficients for the signals are highly significant ($p = 0.000$) and 50% and 30% greater in magnitude in comparison with the nonparametric estimates for one and two higher signals, respectively. These observations suggest that recent signals dominate accumulated signals. But experience likely matters. Subjects, for example, could observe that they were losing auctions despite their bidding the lowest.

Tables 10 and 11 report the intensive margins of learning effects as represented by changes in bid/cost. On average, subjects who saw two higher signals in the previous auction *and* increase their bid/cost do so by 0.05, which is larger than the 0.04 increase after they saw only one higher signal ($p = 0.116$), and larger than the 0.03 increase after they saw no higher signal (i.e., two lower signals). Also, subjects who saw two higher signals and decrease their bid/cost do so by a smaller amount: 0.03 versus 0.04 when they saw only one higher signal ($p = 0.075$). Many of these magnitudes change significantly when I consider only bids over 420. The "Away From 475" magnitudes decrease perhaps because the restriction weeds out bids that fall below 420 after being very high in the previous period. I again wonder whether experience is driving these results, so I reproduce the tables for only periods 1-20 and 21-40. There are only two statistically significant differences between the halves.[37]

Figure 10 shows, by session, how the mean bid changes across periods. Sessions two and three show the most increase; session four and six show the least, though the latter starts out high. Session three even shows a slight downward trend, though the mean of period 40 is higher than that of period one. Interestingly, three sessions — one, two and six — end with nearly identical means, as do sessions four and five, even though each session of these two groups start out with very different means. Learning may partly explain both the bid increases and drop in variance.

Of course, subjects also can learn to act unprofitably. The post-session survey asks, "If you did not submit an initial offer of 475 in every period, why didn't you?" One subject answers, "My competitors were submitting very low prices." More than one subject mentions using the selling price as a guide, with one subject writing simply, "I decided my initial offer price by what

[37] For both all bids and bids above 420, subjects who see no higher signals in the second half raise their bid/cost by 0.04, compared with 0.03 in the first half.

the previous one sold for." Subjects see only selling prices that are at most the value of their own bid, so this strategy could have trapped some subjects in unprofitable behavior.

2.4 Conclusion

Compte et al. prove the general existence of a collusive equilibrium in which all sellers bid the government's reserve and submit the highest bribe the agent will accept. In the lab I implement a specific version of their theory to test the extent to which the equilibrium emerges in a simplified setting. Only 13.5% of auctions in the closing periods display the perfectly collusive outcome because only half of subjects attempt such collusion. Yet due to imperfectly collusive bidding, another 45.2% of selling prices are above 420 but below 475. Therefore in most auctions the selling price is much higher and efficiency much lower than would be expected if subjects simply raised their honest-auction bids by the maximum bribe. Much of this paper has tried to explain why many subjects, having chosen to bid above 420, fail to realize that the most sensible bid is 475. Confusion and gender seem to be the main reasons, though some subjects learn from others. My paper, while contributing to the corruption literature, holds a firmer place in the experimental literature on collusion, in which little evidence exists of tacit collusion among sellers who are continually rematched. An important difference between my study and previous auction studies is that bounded rationality, much more than trust, is the limiting factor for collusion.

My existence proof of Compte et al.'s collusive equilibrium in the lab does not determine whether the theory is applicable to more complicated markets. Even so, I argue that real-world firms have a greater chance at optimal collusion. Especially in the world of lucrative government contracts, corrupt firms probably often do wish to pay a bribe beyond what the agent will accept, and this market imperfection probably does often lead to collusion, even if such collusion better resembles one of the more complicated models described by Compte et al. Furthermore, firms are less likely than lab subjects to face hindrances such as confusion and joy of winning (which are not accounted for in Compte et al.'s model anyhow), and thus may achieve collusion more easily. Firms also may choose explicit collusion over tacit collusion due to the latter's fragility. One possibility for future research is allowing subjects to communicate before each period, which would likely lead to less confusion and more perfect collusion, albeit explicit.

Other research ideas abound, beginning with the study of four sellers instead of three. Also, Compte et al.'s theory doesn't preclude fixed matching instead of random matching. Suppliers in the field likely encounter the same firms again and again. As mentioned earlier, fixed matching should give collusion a better shot, with the side benefit of greater statistical power. Firm decisions also are often made by groups, and in the lab two brains (or more) may be better than one in finding the perfectly collusive equilibrium. Limited bribery is the crux of the equilibrium, so an obvious treatment idea is to remove the bribe ceiling. Implementation, however, might be difficult. Compte et al. also show that increasing control over the agent, i.e., reducing the upper limit for bribes, can increase the ability of firms to collude. So another treatment could be to lower the bribe limit or, equivalently, adjust the cost spread. Alternatively it might be interesting to lower the reserve to, say, 430. Subjects trying to collude might benefit from the strategy-space reduction, and I could more properly incentivize competitive behavior. Due to the current low exchange rate, one could argue that sensible subjects would bid 475 even when other subjects attempt to knock them out. This reduces the strategic nature of the setting.

Finally, Fonseca and Normann (2012) study experimental Bertrand oligopolies in which subjects can communicate. The authors find that subjects continue to collude even after communication is disabled. In a similar fashion, if a design change could foster substantially more reserve-price collusion, a further treatment could be to disable bribery after a certain period to see if collusion persists.

References

Arozamena, Leandro, and Federico Weinschelbaum. “The effect of corruption on bidding behavior in first-price auctions.” *European Economic Review* 53.6 (2009): 645-657.

Beck, Paul J., and Michael W. Maher. “A comparison of bribery and bidding in thin markets.” *Economics Letters* 20.1 (1986): 1-5.

Büchner, Susanne, et al. “Bribery and public procurement: an experimental study.” *Public Choice* 137.1-2 (2008): 103-117.

Burguet, Roberto, and Yeon-Koo Che. “Competitive procurement with corruption.” *RAND Journal of Economics* (2004): 50-68.

Carpineti, Laura, Gustavo Piga, and Matteo Zanza. “The variety of procurement practice: evidence from public procurement.” *Handbook of Procurement*. Cambridge University Press (2006).

Chen, Yan, Peter Katuščák, and Emre Ozdenoren. “Why can’t a woman bid more like a man?” *Games and Economic Behavior* 77.1 (2013): 181-213.

Compte, Olivier, Ariane Lambert-Mogiliansky, and Thierry Verdier. “Corruption and competition in procurement auctions.” *Rand Journal of Economics* (2005): 1-15.

Crawford, Vincent P. “Boundedly rational versus optimization-based models of strategic thinking and learning in games.” *Journal of Economic Literature* 51.2 (2013): 512-527.

Croson, Rachel, and Uri Gneezy. “Gender differences in preferences.” *Journal of Economic literature* 47.2 (2009): 448-474.

Engelbrecht-Wiggans, Richard, Ernan Haruvy, and Elena Katok. “A comparison of buyer-determined and price-based multiattribute mechanisms.” *Marketing Science* 26.5 (2007): 629-641.

Fischbacher, Urs. “z-Tree: Zurich toolbox for ready-made economic experiments.” *Experimental economics* 10.2 (2007): 171-178.

Fonseca, Miguel A., and Hans-Theo Normann. “Explicit vs. tacit collusion — The impact of communication in oligopoly experiments.” *European Economic Review* 56.8 (2012): 1759-1772.

Fugger, Nicolas, Elena Katok, and Achim Wambach. “Collusion in dynamic buyer-determined reverse auctions.” *Management Science* 62.2 (2015): 518-533.

Ham, John C., and John H. Kagel. “Gender effects in private value auctions.” *Economics Letters* 92.3 (2006): 375-382.

Hao, Li, and Daniel Houser. “Perceptions, intentions, and cheating.” *Journal of Economic Behavior & Organization* 133 (2017): 52-73.

Kagel, J. H. “Auctions: a survey of experimental research,” in (J. Kagel and AE Roth, eds.), *The Handbook of Experimental Economics* pp. 501-85.” (1995).

Kagel, John H., and Dan Levin. “Auctions: A Survey of Experimental Research.” Mimeo. (2014)

Lambert-Mogiliansky, Ariane. “Corruption and collusion: strategic complements in procurement.” *International Handbook on the Economics of Corruption* 2 (2011): 108-140.

Lien, Da-Hsiang Donald. “A note on competitive bribery games.” *Economics Letters* 22.4 (1986): 337-341.

Lynn, Michael, Sean Masaki Flynn, and Chelsea Helion. “Do consumers prefer round prices? Evidence from pay-what-you-want decisions and self-pumped gasoline purchases.” *Journal of Economic Psychology* 36 (2013): 96-102.

OECD (2008), *Fighting Cartels in Public Procurement.* OECD. Paris.

OECD (2014), *OECD Foreign Bribery Report: Analysis of the Crime of Bribery of Foreign Public Officials,* OECD Publishing. http://dx.doi.org/10.1787/9789264226616-en

Selten, Reinhard. “12 Learning direction theory and impulse balance equilibrium.” *Economics Lab* (2004): 133.

Toussaint, Loren, and Jon R. Webb. “Gender differences in the relationship between empathy and forgiveness.” *The Journal of Social Psychology* 145.6 (2005): 673-685.

Appendix

Tables and Figures

Table 1: Summary of Experiment

No. of sessions: 6, all identical. *No. of subjects per session:* 9.
Exchange rate: E$40.00 = U.S.$1.00.
Mean earnings in U.S. dollars: $13.76, including a $5 show-up fee

Note: I report on one additional session of auctions without bribery, in the appendix.

Table 2: Summary of Nomenclature

Types of Auction
Corrupt: the auction of the experiment (with bribery).
Honest: the auction of the experiment minus the bribery stage.
Competitive: corrupt auction in which all subjects add 20 to the risk-neutral Nash equilibrium (RNNE) bid they would have submitted in an honest auction.
Imperfectly collusive: corrupt auction with selling price > 420, < 475 (because at least one subject bids in this region).
Perfectly collusive: corrupt auction with selling price = 475 (because all subjects bid 475).
Types of Bid
Competitive: bid ≤ 420 (because the highest competitive bid is 420).
Imperfectly collusive: bid > 420, < 475.
Perfectly collusive: bid = 475.

Notes: I report on one session of honest auctions, in the Appendix. I use hypothetical outcomes of honest auctions, not experimental results, for comparisons. For competitive auctions, selling price ≤ 420, but the price also can be in this region when at least one subject bids competitively.

Table 3: Means and Predictions for Bid, Selling Price, Subject Profit, and Efficiency

	Competitive Prediction		*Bid*		*Perfectly Collusive Prediction*
Periods 1-7	411.07	$<^{**}$	444.96 (12.29)	$<^{**}$	475.00
			$<^{**}$		
Periods 34-40	412.05	$<^{**}$	455.30 (8.98)	$<^{**}$	475.00
			Selling Price		
Periods 1-7	406.22	$<^{**}$	423.82 (14.90)	$<^{**}$	475.00
			$\approx^{\dagger}$		
Periods 34-40	407.74	$<^{**}$	433.13 (12.50)	$<^{**}$	475.00
			Subject Profit		
Periods 1-7	2.08 (0.09)	$<^{**}$	7.17 (4.08)	$<^{**}$	22.77 (0.48)
			$\approx^{\dagger\dagger}$		
Periods 34-40	1.66 (0.14)	$<^{**}$	9.36 (3.83)	$<^{**}$	22.19 (0.38)
			Efficiency		
Periods 1-7	0.73 (0.06)	$>^{**}$	0.48 (0.11)	$>^{**}$	0.30 (0.12)
			$\approx$		
Periods 34-40	0.71 (0.04)	$>^{**}$	0.46 (0.13)	$>^{**}$	0.28 (0.09)

Notes: The unit of observation is the session. Standard deviations are in parentheses. Competitive bid predictions are transformations, shifted upward by 20, of the cost draws using the Nash equilibrium bidding function for risk-neutral sellers who compete in honest auctions with the same relevant parameters. Competitive bidders submit the highest bribe they can afford. Perfectly collusive predictions are simulations in which subjects, with the same cost draws, bid 475 and submit 20 as their bribe. Cost draws were the same in each session. ** indicates a significant difference at the 5% level, according to the Wilcoxon signed-rank test. † p = 0.116 †† p = 0.173.

Table 4: Bids and Selling Prices by Region for Periods 34-40

	% ≤ 420	*% > 420, ≠ 475*	*% = 475*
Bids	16.4	34.4	49.2
Selling Prices	40.5	45.2	14.3

Table 5: Means and Predictions for Fee Offer and Winning Fee

	Fee Offer		*Prediction*
Periods 1-7	15.04 (2.88)	$<^{**}$	18.16 (1.67)
	$<^{**}$		≈
Periods 34-40	16.98 (2.10)	$<^{**}$	17.97 (1.67)
	Winning Fee		
Periods 1-7	18.49 (1.57)	$\approx^{\dagger}$	19.18 (0.99)
	≈		≈
Periods 34-40	18.58 (1.80)	$<^{*}$	18.97 (1.26)

Notes: The unit of observation is the session. Standard deviations are in parentheses. Predictions are the highest fees subjects could afford given the minimum first-stage bid. ** and * indicate a significant difference at the 5% and 10% levels, respectively, according to the Wilcoxon signed-rank test. † p = 0.140

Table 6: Proportion of Bids Classified as Competitive (≤ 420), Imperfectly Collusive (> 420, ≠ 475) and Perfectly Collusive (= 475) by Gender and Confusion

		Confusion			*Gender*			*Confusion | Male*			*Confusion | Female*		
Bid Values	*All*	*Confused*		*Unconfused*	*Male*		*Female*	*Confused Male*		*Unconfused Male*	*Confused Female*		*Unconfused Female*
≤ 420	0.18	0.26	$>^{*}$	0.11	0.16	$\approx$	0.19	0.24	$\approx^{\dagger}$	0.09	0.29	$>^{*}$	0.11
	(0.08)	(0.15)		(0.09)	(0.08)		(0.10)	(0.16)		(0.10)	(0.24)		(0.07)
	$<^{**}$	$<^{*}$		$<^{**}$	$<^{**}$		$<^{**}$	$<^{*}$		$<^{**}$	$\approx^{\dagger}$		$<^{**}$
> 420, ≠ 475	0.42	0.48	$\approx$	0.37	0.29	$<^{**}$	0.53	0.39	$>^{*}$	0.21	0.55	$\approx$	0.46
	(0.12)	(0.14)		(0.17)	(0.12)		(0.17)	(0.10)		(0.18)	(0.19)		(0.25)
	$\approx$	$\approx^{\dagger}$		$\approx$	$<^{*}$		$\approx$	$\approx$		$<^{*}$	$>^{*}$		$\approx$
= 475	0.40	0.26	$<^{**}$	0.51	0.54	$>^{**}$	0.29	0.37	$<^{**}$	0.70	0.16	$<^{**}$	0.44
	(0.19)	(0.18)		(0.23)	(0.16)		(0.26)	(0.19)		(0.28)	(0.16)		(0.28)

Notes: “Confused” indicates that a subject missed four or more questions on the comprehension test. The unit of observation is the session. Standard deviations are in parentheses. ** and * indicate a significant difference at the 5% and 10% levels, respectively, according to the Wilcoxon signed-rank test. †p = 0.116.

Table 7: Marginal Effects From Tobit Regressions With Random Effects

Dependent Variable: Bid	*All Data*		*Periods 1-7*		*Periods 34-40*		*Bids > 420*		*Bids ≤ 420*	
Cost	0.335***	0.310***	0.430***	0.467***	0.251***	0.264***	0.098**	0.066*	0.436***	0.449***
	(0.051)	(0.050)	(0.115)	(0.116)	(0.086)	(0.087)	(0.039)	(0.036)	(0.059)	(0.059)
Period		0.301***		1.82***		0.506*		0.300***		-0.061*
		(0.035)		(0.415)		(0.299)		(0.033)		(0.034)
Male		8.75*		4.65		12.26**		8.25**		-3.97*
		(4.95)		(5.14)		(5.68)		(3.68)		(2.14)
Confusion ^{a}		-15.5***		-18.79***		-10.33*		-10.81***		2.54
		(5.00)		(5.15)		(5.59)		(3.64)		(2.10)
Log Likelihood	-6,402	-6,342	-1,345	-1,328	-965	-959	-4,297	-4,211	-1,351	-1,347
Observations (Groups)	2,160 (54)		378 (54)		378 (54)		1,772 (54)		388 (41)	

Notes: Regressions also include a set of session dummies. Delta-method standard errors are in parentheses. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively. a) “Confusion” indicates that a subject missed four or more questions on the comprehension test.

Table 8: Probabilities That Subjects Raise or Lower Their Bids After Observing Their Competitors' Bids in the Previous Auction

Trajectory (Subject did not bid 475 in previous period)	*2 Lower Signals*		*1 Higher Signal*		*2 Higher Signals*
Toward 475	0.27 (0.14)	$<^{*}$	0.41 (0.09)	$<^{**}$	0.59 (0.07)
	$<^{*}$		$\approx$		$>^{**}$
Away From 475	0.43 (0.09)	$\approx$	0.40 (0.05)	$>^{**}$	0.29 (0.07)
	$>^{*}$		$>^{**}$		$>^{**}$
Unchanged	0.30 (0.14)	$>^{**}$	0.19 (0.10)	$>^{**}$	0.12 (0.08)
Trajectory (Subject bid 475 in previous period)	*0 Lower Signal*		*1 Lower Signal*		*2 Lower Signals*
Away From 475	0.09 (0.05)	$\approx$	0.13 (0.06)	$\approx$	0.10 (0.03)

Table 9: Re-creation of Table 7 Using Only Bids Greater Than 420, With Indicators of Significant Differences From Respective Values of Table 7

Trajectory (Subject did not bid 475 in previous period)	*2 Lower Signals*		*1 Higher Signal*		*2 Higher Signals*
Toward 475	0.32^{**} (0.17)	$<^{**}$	0.50^{**} (0.10)	$<^{**}$	0.71^{**} (0.05)
	$\approx$		$>^{**}$		$>^{**}$
Away From 475	0.34^{**} (0.12)	$\approx$	0.28^{**} (0.05)	$>^{**}$	0.18^{**} (0.05)
	$\approx$		$\approx$		$\approx$
Unchanged	0.34^{**} (0.13)	$>^{**}$	0.21^{**} (0.11)	$>^{**}$	0.12 (0.09)
Trajectory (Subject bid 475 in previous period)	*0 Lower Signal*		*1 Lower Signal*		*2 Lower Signals*
Away From 475	0.07 (0.02)	$\approx$	0.12^{*} (0.06)	$\approx^{\dagger}$	0.07^{*} (0.02)

Notes: A higher (lower) signal is a competitor's bid in the previous auction that was higher (lower) than the subject's own bid. The unit of observation is the session. Standard deviations are in parentheses. ** and * indicate a significant difference at the 5% and 10% levels, respectively, according to the Wilcoxon signed-rank test. † $p = 0.116$.

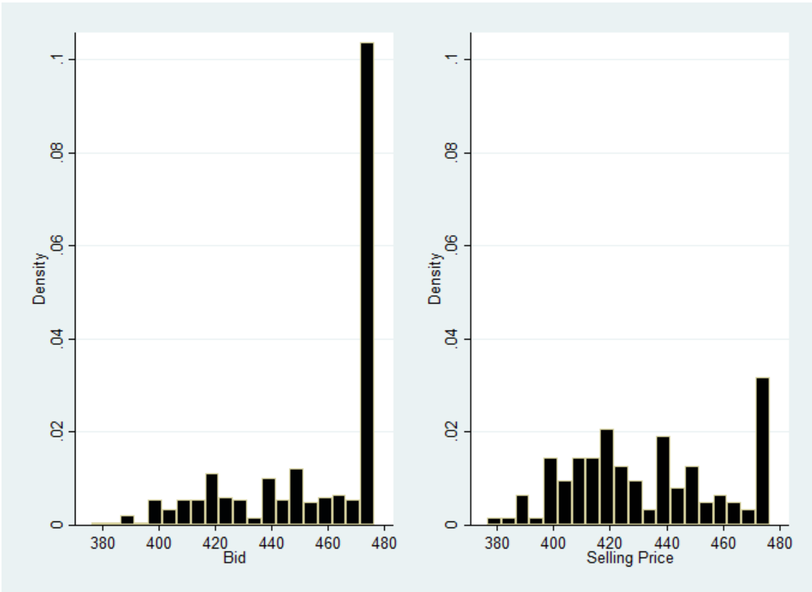
Table 10: Mean Change in Bid/Cost by Which Subjects Raise or Lower Their Bids After Observing Their Competitors' Bids in the Previous Auction, Conditional on Subjects Raising or Lowering

Trajectory (Subject did not bid 475 in previous period)	*2 Lower Signals*		*1 Higher Signal*		*2 Higher Signals*
Toward 475	0.032 (0.008)	$\approx$	0.039 (0.012)	$\approx^{\dagger}$	0.051 (0.011)
	$<^{*}$		$\approx$		$>^{**}$
Away From 475	0.039 (0.012)	$\approx$	0.041 (0.015)	$>^{*}$	0.028 (0.007)
Trajectory (Subject bid 475 in previous period)	*0 Lower Signal*		*1 Lower Signal*		*2 Lower Signals*
Away From 475	0.059 (0.032)	$\approx$	0.084 (0.013)	$\approx$	0.083 (0.030)

Table 11: Re-creation of Table 10 Using Only Bids Greater Than 420, With Indicators of Significant Differences From Respective Values of Table 6

Trajectory (Subject did not bid 475 in previous period)	*2 Lower Signals*		*1 Higher Signal*		*2 Higher Signals*
Toward 475	0.032 (0.008)	$\approx$	0.040^{*} (0.012)	$<^{**}$	0.059^{**} (0.014)
	$\approx$		$\approx^{\dagger}$		$>^{**}$
Away From 475	0.030^{**} (0.007)	$\approx$	$0.031^{\dagger}$ (0.008)	$\approx$	0.025 (0.006)
Trajectory (Subject bid 475 in previous period)	*0 Lower Signal*		*1 Lower Signal*		*2 Lower Signals*
Away From 475	0.043 (0.043)	$<^{*}$	0.069^{*} (0.012)	$>^{**}$	0.051^{*} (0.012)

Notes: A higher (lower) signal is a competitor's bid in the previous auction that was higher (lower) than the subject's own bid. The unit of observation is the session. Standard deviations are in parentheses. ** and * indicate a significant difference at the 5% and 10% levels, respectively, according to the Wilcoxon signed-rank test. † $p = 0.116$.



Note: Bin width = 5.00. Figure A1 in the Appendix shows the same histograms with bin width = 1.00.

Figure 1: Bids and Selling Prices, Periods 34-40

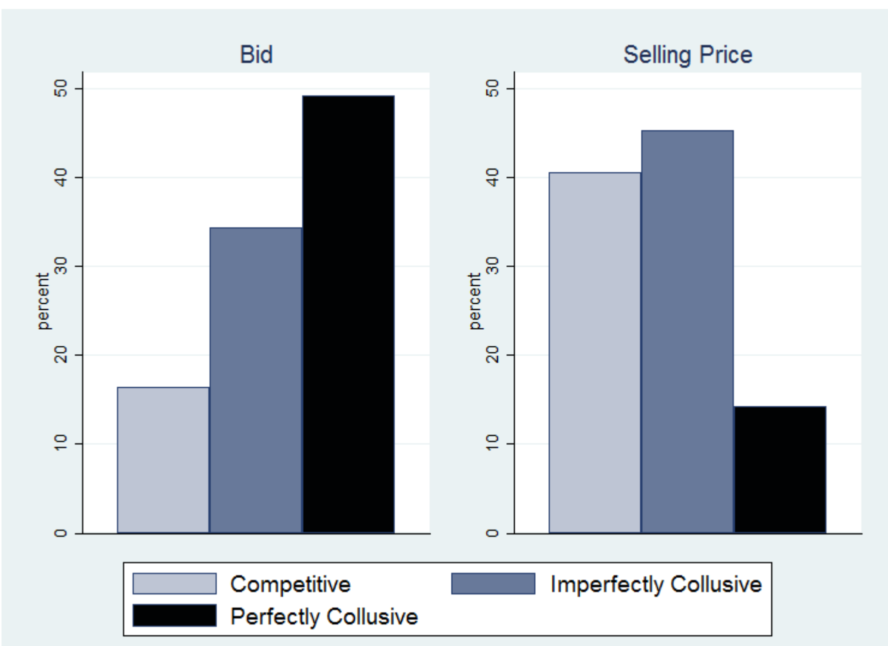


Figure 2: Bids and Selling Prices, Periods 34-40

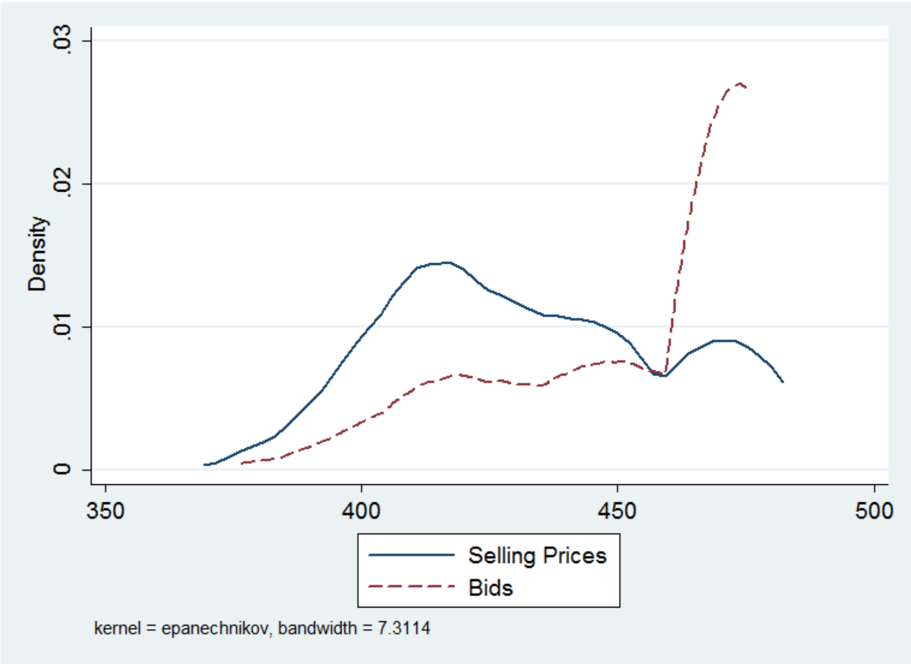
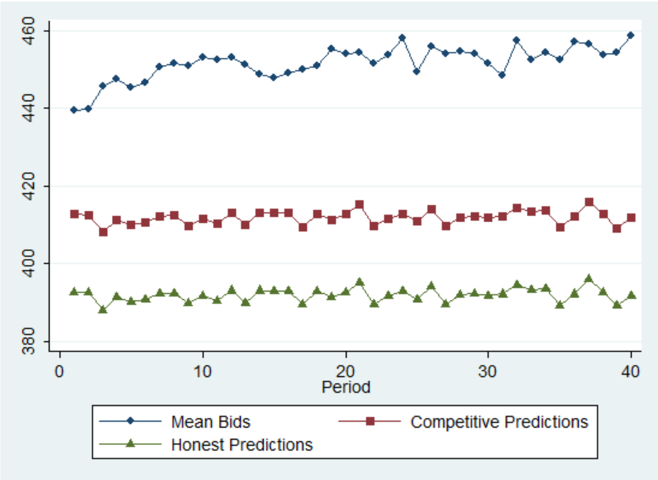
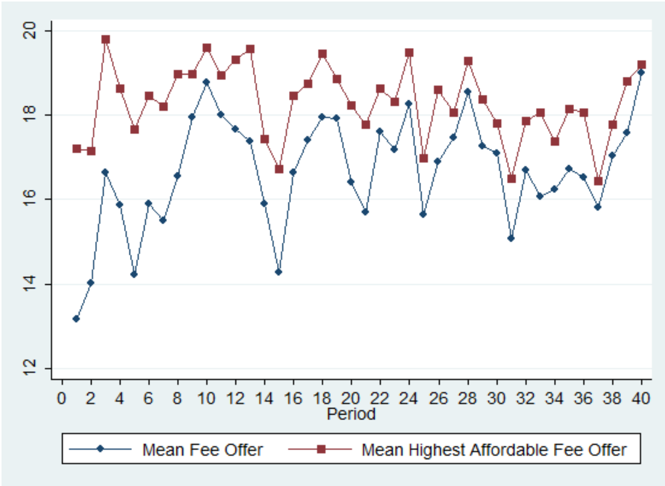


Figure 3: Kernel Densities of Selling Prices and Bids, Periods 34-40



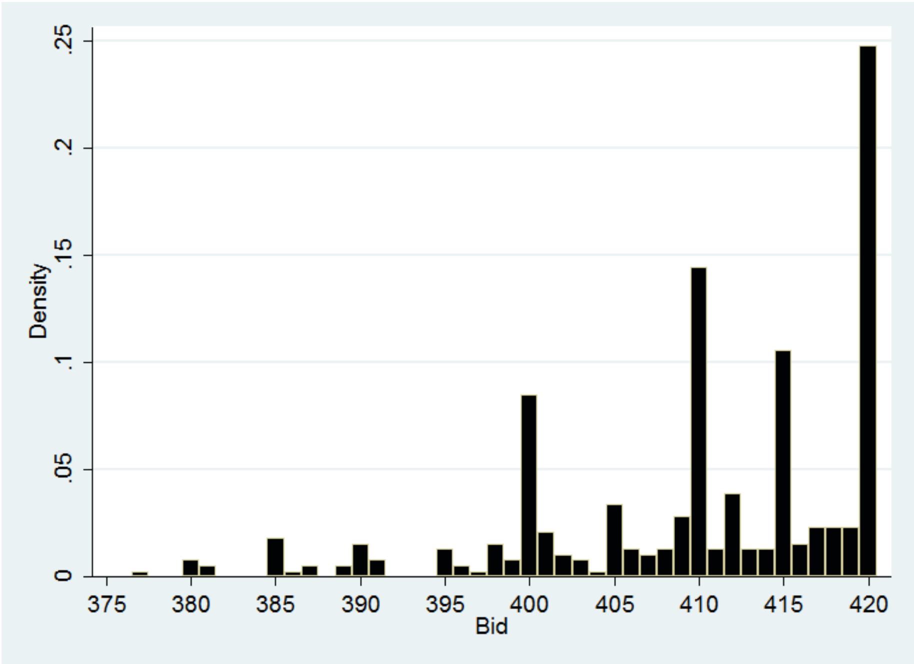
Notes: Honest predictions are transformations of the cost draws using the Nash equilibrium bidding function for risk-neutral sellers who compete in auctions identical to those of the experiment but without bribery. Competitive predictions are the transformations of the cost draws using the same bidding function shifted upward by 20.

Figure 4: Mean Bids and Predictions by Period



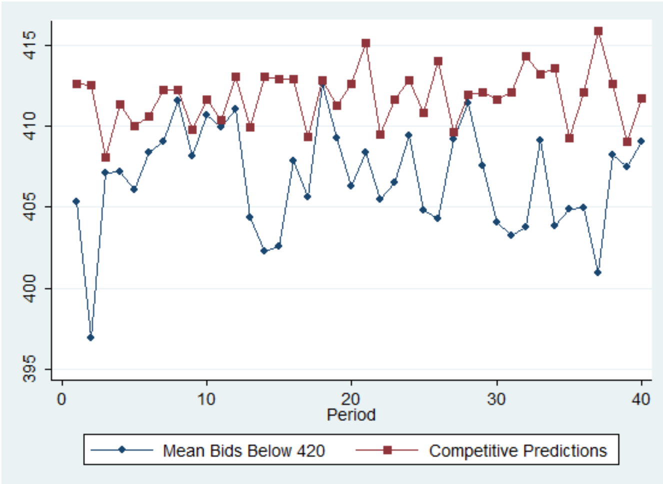
Note: The highest affordable fee offer for each subject is whichever is highest: 20 or the difference between cost and the lowest bid.

Figure 5: Mean Fee Offers: Actual vs. Highest Subjects Could Afford

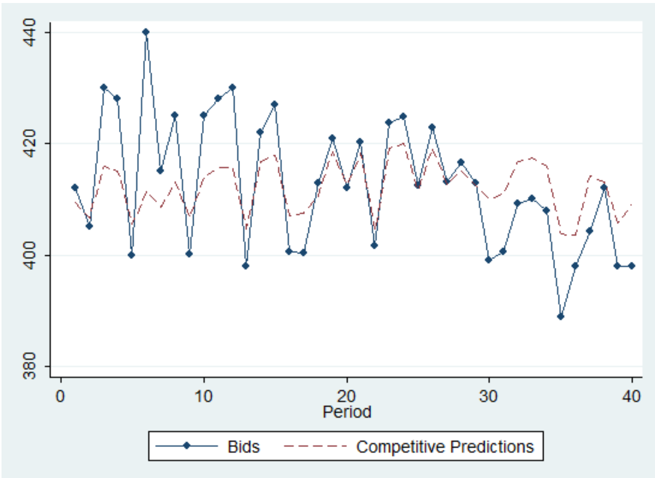


Notes: Bin width = 1.00. Histogram excludes one bid of 304.76.

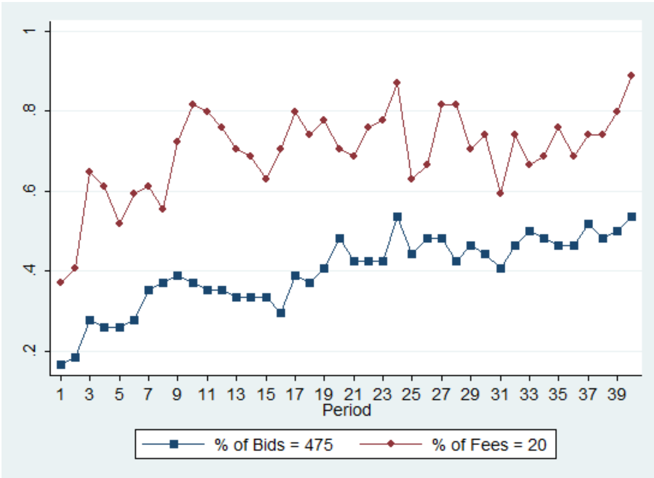
Figure 6: Bids of 420 and Below



Notes: Competitive predictions are transformations of the cost draws using the bidding function for honest auctions shifted upward by 20. Cost draws were the same in each session.
Figure 7: Mean Bids Below 420 and Predictions by Period



Notes: Competitive predictions are transformations of the cost draws using the bidding function for honest auctions shifted upward by 20. Cost draws were the same in each session.
Figure 8: Bids of Subject 3, Session 5



	% of Bids = 475	% of Fees = 20
Periods 1-7	0.25	0.54
	(0.17)	(0.24)
	<**	<**
Periods 34-40	0.49	0.76
	(0.21)	(0.17)

Notes: The unit of observation is the session. Standard deviations are in parentheses. ** indicates a significant difference at the 5% level according to the Wilcoxon signed-rank test.

Figure 9: **Extreme Values as a % of Bids and Fees**

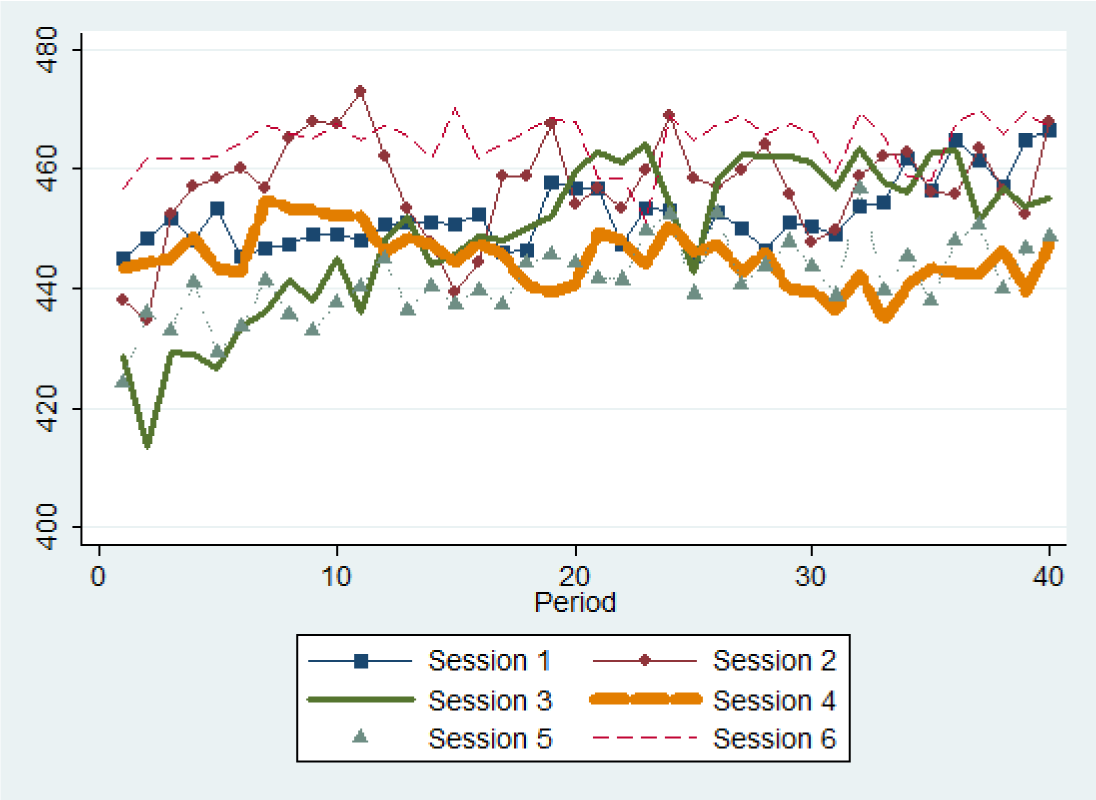
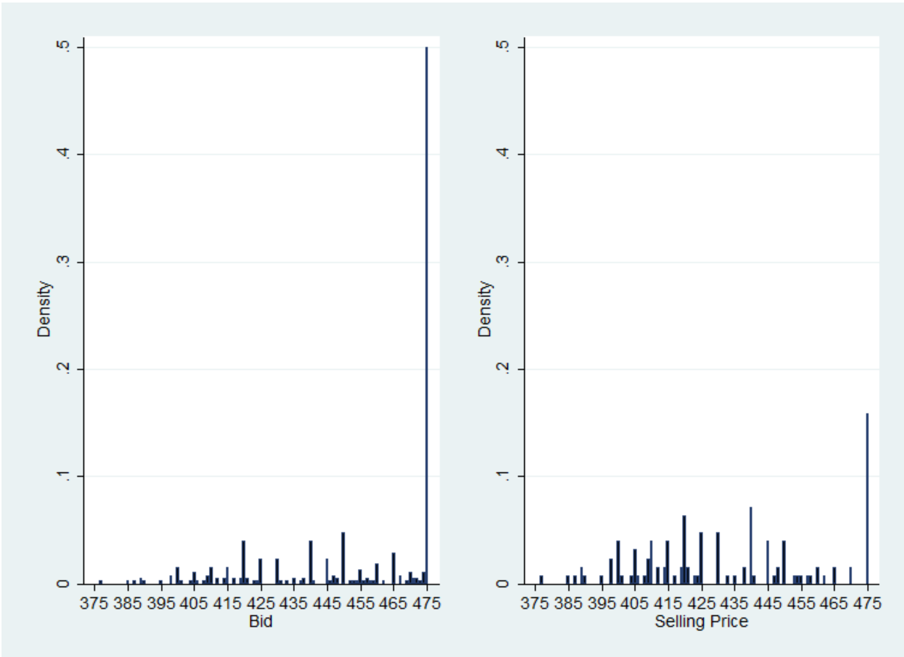


Figure 10: Change in Mean Bids by Session

Additional Tables and Figures, and Experimental Instructions



Note: Bin width = 1.00.
Figure A1: Bids and Selling Prices, Periods 34-40

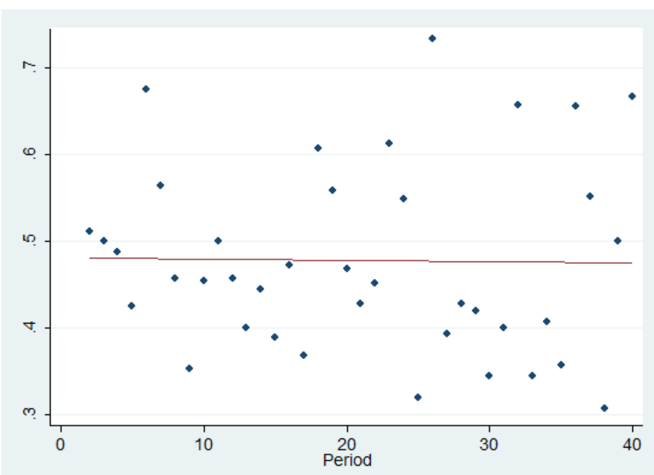


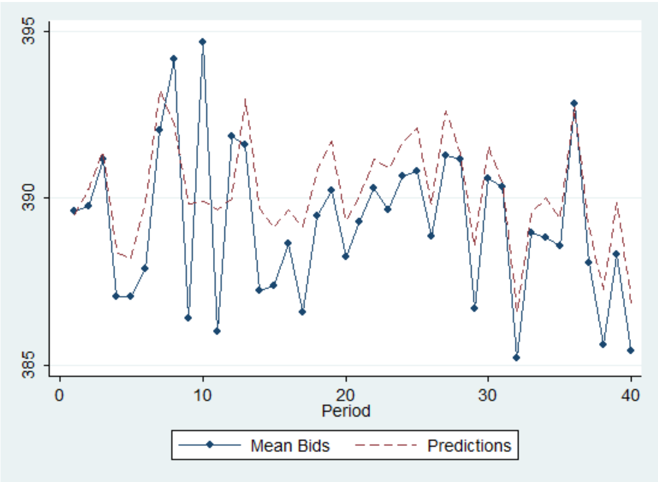
Figure A2: % of Bids in the Direction of 475, Assuming Subject Did Not Bid 475 in Previous Period

‘Honest’ Session Results

In addition to the six sessions described in the main text, I conducted one session that lacked the bribery stage. That is, subjects participated in “honest” procurement auctions. The only other difference between this session and the others is that subjects faced a different sequence of costs, drawn from $U[370.00, 400.00]$. The instructions, however, stated that costs were drawn from $U[375.00, 400.00]$, as in the corrupt sessions. Due to a typo in the computer program, subjects actually faced the wider distribution. Earnings were given in experimental dollars (E$) with an exchange rate of E$10.00 = U.S.$1.00. Subjects who participated in the 45-minute session earned on average U.S.$10.53, including a U.S.$5 show-up fee. Henceforth all monetary figures are in E$.

Results mostly match simulated predictions for the one session with honest auctions (Figure A2). The mean winning price is 382.90, slightly below the predicted value of 385.40. The mean bid is 389.21, very close to the predicted 390.17. The mean profit is 1.38, well below the predicted 2.43. The efficiency rate is 0.85. Spearman correlation between cost and offer price is 0.88, rising to 0.94 after the removal from the data set of one of the nine subjects. Both coefficients are highly significant.

Besides supporting theory, these results are evidence that subjects are capable of understanding the rules of a procurement auction. Despite the error in the computer program mentioned earlier, I did not choose to rerun the session, or to run more sessions. The results speak for themselves, and previous experimental research already has shown little evidence of collusion among three sellers who are rematched and cannot communicate.



Notes: Honest predictions are transformations of the cost draws using the Nash equilibrium bidding function for risk-neutral sellers: $bid = (2/3) \times cost + 400/3$.

Figure A3: Mean Bids for Session With Honest Auctions

Comparisons Among Tables of Means Using Different Period Blocks

I reproduce Table 3 using different blocks of periods other than the first seven periods and last seven periods used in main text, specifically the first and last five periods as well as the first and last 10 periods (Tables A2 and A3). I chose the first and last seven periods because between these periods is when the mean bid becomes significantly different. So I expected statistical differences among these tables, even though the point estimates are similar.

Below is a comparison of the statistical significance levels among the tables. First consider the differences between the beginning and closing periods. As expected, for the comparison between the first and last 10 periods, the mean bid is statistically different at only the 11.6% level instead of at the 5% level for the first and last seven periods. All other comparisons in regard to the first and last 10 periods are even less significant. For the comparison between the first and last five periods, the selling price is now significantly different at the 7.47% level, compared with only at the 11.6% level when comparing the first and last seven periods. There are no other differences between the seven-period tables and five-period tables.

Next consider the differences between the outcomes and the predictions, all of which are significant at the 5% level for the seven-period tables and ten-period tables. For the five-period tables, all are significant at the 5% level except for two. The selling price for periods 1-5 is significant from the competitive prediction at only the 7.47% level. Efficiency for periods 36-40 is significant from the competitive prediction at only the 14% level.

Now consider the differences in the point estimates among the table entries, all at the 5% level. The mean bids for periods 1-5, 1-7, and 1-10 are all distinguishable from each other; the selling price for periods 1-5 is statistically different form the selling price for periods 1-10; and the subject profit for periods 1-7 is distinguishable from the subject profit for periods 1-10.

Table A1: Means and Predictions for Bid, Selling Price, Subject Profit, and Efficiency

	Competitive Prediction		*Bid*		*Perfectly Collusive Prediction*
Periods 1-5	410.93	$<^{**}$	443.51 (12.69)	$<^{**}$	475.00
			$<^{**}$		
Periods 36-40	412.30	$<^{**}$	456.09 (9.30)	$<^{**}$	475.00
			Selling Price		
Periods 1-5	406.30	$<^{*\dagger}$	423.15 (15.27)	$<^{**}$	475.00
			$<^{*\dagger}$		
Periods 36-40	408.37	$<^{**}$	432.81 (14.57)	$<^{**}$	475.00
			Subject Profit		
Periods 1-5	1.99 (0.09)	$<^{**}$	7.17 (4.27)	$<^{**}$	22.72 (0.33)
			$\approx^{\dagger\dagger}$		
Periods 36-40	1.47 (0.14)	$<^{**}$	9.21 (4.42)	$<^{**}$	22.20 (0.26)
			Efficiency		
Periods 1-5	0.70 (0.06)	$>^{**}$	0.50 (0.14)	$>^{**}$	0.30 (0.07)
			$\approx$		
Periods 36-40	0.62 (0.11)	$\approx^{\dagger\dagger\dagger}$	0.48 (0.18)	$>^{**}$	0.31 (0.05)

Notes: The unit of observation is the session. Standard deviations are in parentheses. Competitive bid predictions are transformations, shifted upward by 20, of the cost draws using the Nash equilibrium bidding function for risk-neutral sellers who compete in honest auctions with the same relevant parameters. Competitive bidders submit the highest bribe they can afford. Perfectly collusive predictions are simulations in which subjects, with the same cost draws, bid 475 and submit 20 as their bribe. Cost draws were the same in each session. ** indicates a significant difference at the 5% level, according to the Wilcoxon signed-rank test. † p = 0.075 †† p = 0.173. ††† p = 0.140.

Table A2: Means and Predictions for Bid, Selling Price, Subject Profit, and Efficiency

	Competitive Prediction		*Bid*		*Perfectly Collusive Prediction*
Periods 1-10	411.12	$<^{**}$	447.02 (12.09)	$<^{**}$	475.00
			$\approx^{\dagger}$		
Periods 31-40	412.41	$<^{**}$	454.52 (9.06)	$<^{**}$	475.00
			Selling Price		
Periods 1-10	406.46	$<^{**}$	426.76 (14.81)	$<^{**}$	475.00
			$\approx$		
Periods 31-40	408.09	$<^{**}$	431.80 (12.59)	$<^{**}$	475.00
			Subject Profit		
Periods 1-10	2.03 (0.10)	$<^{**}$	8.02 (4.31)	$<^{**}$	22.63 (0.34)
			$\approx$		
Periods 31-40	1.69 (0.11)	$<^{**}$	8.83 (3.48)	$<^{**}$	22.29 (0.38)
			Efficiency		
Periods 1-10	0.74 (0.06)	$>^{**}$	0.51 (0.10)	$>^{**}$	0.28 (0.08)
			$\approx$		
Periods 31-40	0.72 (0.08)	$>^{**}$	0.47 (0.15)	$\approx$	0.36 (0.10)

Notes: The unit of observation is the session. Standard deviations are in parentheses. Competitive bid predictions are transformations, shifted upward by 20, of the cost draws using the Nash equilibrium bidding function for risk-neutral sellers who compete in honest auctions with the same relevant parameters. Competitive bidders submit the highest bribe they can afford. Perfectly collusive predictions are simulations in which subjects, with the same cost draws, bid 475 and submit 20 as their bribe. Cost draws were the same in each session. ** indicates a significant difference at the 5% level, according to the Wilcoxon signed-rank test. † $p = 0.116$.

Instructions

Welcome! You are being given US$5.00 for participating in this study. Please read these instructions carefully. If you make good decisions today, you can make a considerable amount of money beyond the US$5.00 starting amount. However, any loss you make will be deducted from this amount.

If you have any questions during the experiment, please raise your hand, and I will come to assist you. Otherwise, please do not talk or communicate with anyone.

You will earn experimental dollars (E$) during the experiment. At the end of the experiment, you will be paid in cash according to the total amount you earned, at the rate of 40.00 E$ for one U.S. dollar: E$40.00 = U.S.$1.00. From now on in this experiment, all dollar amounts are in E$.

Overview

There are 40 periods. In each period you are a seller trying to sell an item to a computerized buyer. There are *two* other sellers (human participants) in your market trying to sell the same item. However, only one item can be sold in the market each period.

Random Assignment of Sellers

In each period, two other sellers are randomly assigned to your market from a group of eight participants. This means you will not be in a market with the same sellers every period. Furthermore, no one will ever know the identity of the other sellers in his or her market.

How Cost Is Determined

At the beginning of each period, each seller learns how much it costs him or her to produce the item. A seller's cost is drawn randomly and is equally likely to be any number (to two decimal places) between $375.00 and $400.00, including $375.00 and $400.00. A seller's cost in one period is unrelated to any seller's cost in any period. That is to say, your cost in this period is unrelated to your costs in previous or later periods, and is unrelated to the other sellers' costs.

How the Seller of the Item Is Determined

A two-stage process each period determines who sells that period's item.

Stage 1:

Each seller makes an *initial* offer price to two decimal places.

The computerized buyer is not willing to pay more than $475.00, so sellers cannot ask for more than $475.00.

Stage 2:

After each of the three sellers has submitted an initial offer price, sellers learn the initial offer prices of their two competitors.

Sellers then compete for the ability to match the lowest initial offer price, even if they themselves submitted the lowest initial offer price. They do so by submitting a *resubmission fee* (to two decimal places) between $0.00 and $20.00, including $0.00 and $20.00.

To understand how the computerized buyer selects who sells the item, consider the two possible cases:

Case 1: At least one seller submits a resubmission fee above zero.

In this case, any seller who alone submits the *highest* resubmission fee automatically resubmits a *final* offer price equal to the lowest initial offer price. If there is a tie in resubmission fees, *one* of the tying sellers is randomly selected to resubmit.

Only the seller selected to resubmit actually pays the resubmission fee. And *only* the seller selected to resubmit sells the item at a *selling price* equal to the lowest initial offer price (now the lowest final offer price).

For sellers not selected to resubmit, their initial offer price *is* their final offer price. They do not pay a resubmission fee or sell the item, even if their final offer price also is lowest.

Sellers who submit a resubmission fee of zero are never selected to resubmit; they also never sell the item, given Case 1's assumption that at least one seller submits a resubmission fee above zero.

Note well: Even sellers who submit the lowest initial offer price in Stage 1 compete in Stage 2 to match their initial offer price and sell the item. Multiple sellers may have the lowest final offer price, but only the seller who is selected to resubmit sells the item.

Case 2: All three sellers submit a resubmission fee of zero.

When all three sellers submit a resubmission fee of zero, nobody is selected to resubmit, and everybody's initial offer price *is* their final offer price. In this case, the seller offering the *lowest* final offer price sells the item, with ties determined randomly.

Your Profit

If you do not sell the item, then your profit for the period is **zero**.

If you do sell the item and were selected to resubmit, your profit is the **difference between the selling price and your cost, minus your resubmission fee.** Keep in mind that *it is possible to lose money if you sell at a price lower than the sum of your cost and resubmission fee.* So it is important to pay attention to your cost (which changes each period) and how it relates to the lowest initial offer price (which will be the selling price).

If you do sell the item and were *not* selected to resubmit (because all three sellers submitted a resubmission fee of zero), your profit is **the difference between the selling price and your cost.** In this case, *it is possible to lose money if you sell at a price lower than your cost.*

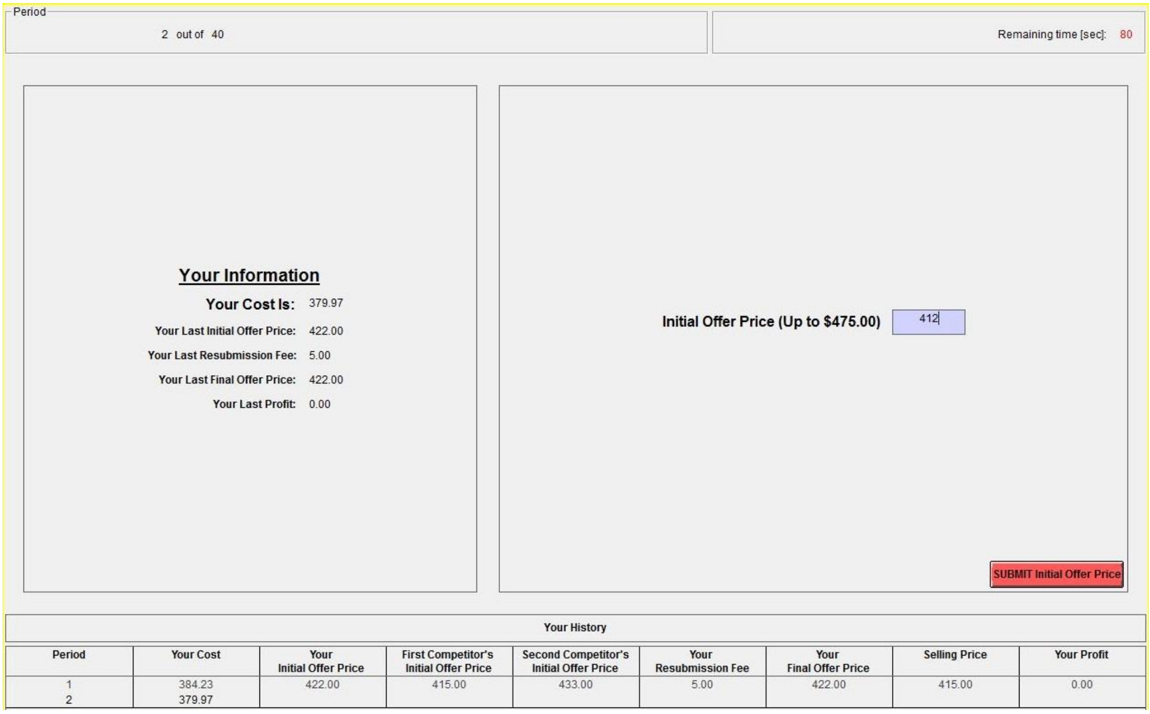
Profit if you do not sell = 0

Profit if you sell = selling price – your cost – your resubmission fee (or, if all sellers submit a resubmission fee of zero, selling price – your cost)

The total amount you earn is the sum of your profits from all periods. Note that if your cumulative profits become negative, you will not be allowed to continue playing.

How to Submit Your Initial Offer Price

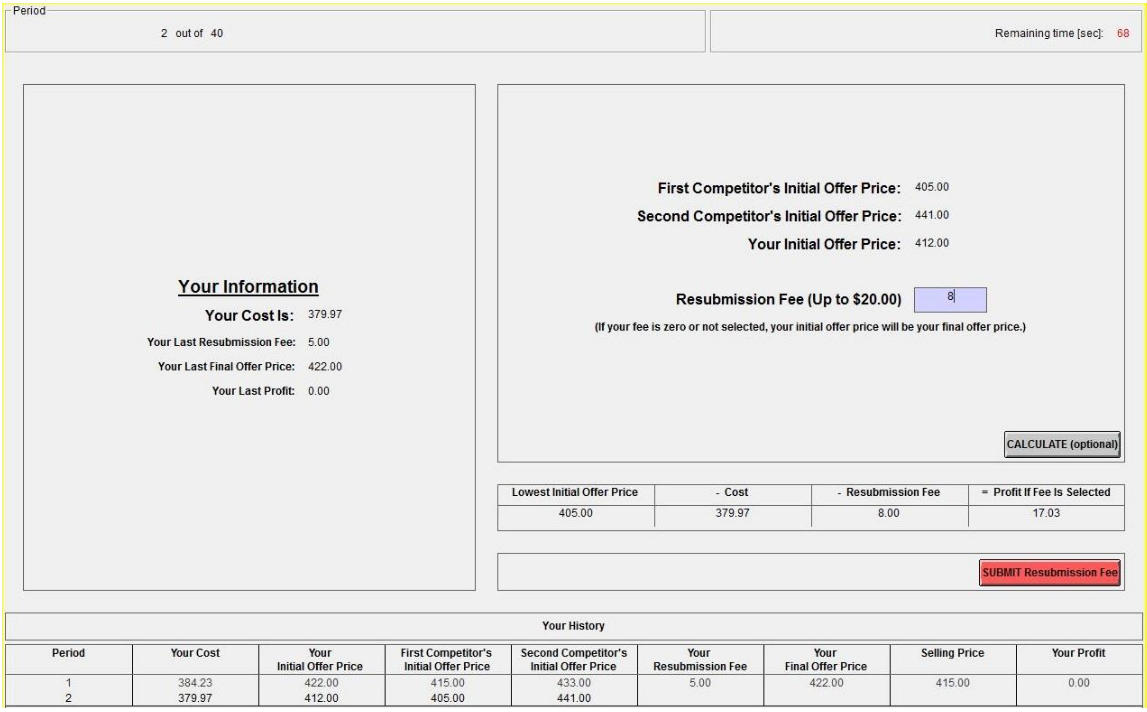
Below is an example of the screen you will be facing. As you can see, your cost is on the left.



To submit an initial offer price, enter your initial offer price amount in the "Initial Offer Price (Up to $475.00)" box and then press the "SUBMIT Initial Offer Price" button.

How to Submit Your Resubmission Fee

Below is an example of the screen you will be facing. As you can see, you are given your cost and (on the right) your initial offer price and the initial offer prices submitted by your two competitors.



Enter your resubmission fee amount in the "Resubmission Fee (Up to $20.00)" box.

You then have the option of using the built-in calculator to view potential profits. The software will display calculations. Let's go over the variables and computation:

1. Lowest Initial Offer Price = This is the lowest initial offer price from Stage 1.

2. Cost = This is your cost.

3. Resubmission Fee = This is the resubmission fee you entered.

4. Profit If Fee Is Selected = lowest initial offer price – cost – resubmission fee. This is your profit if your resubmission fee is selected.

You can use the optional calculator for any number of resubmission fees you wish. When you are ready to submit your resubmission fee, press the "SUBMIT Resubmission Fee" button.

You also can skip doing calculations and immediately click "SUBMIT Resubmission Fee."

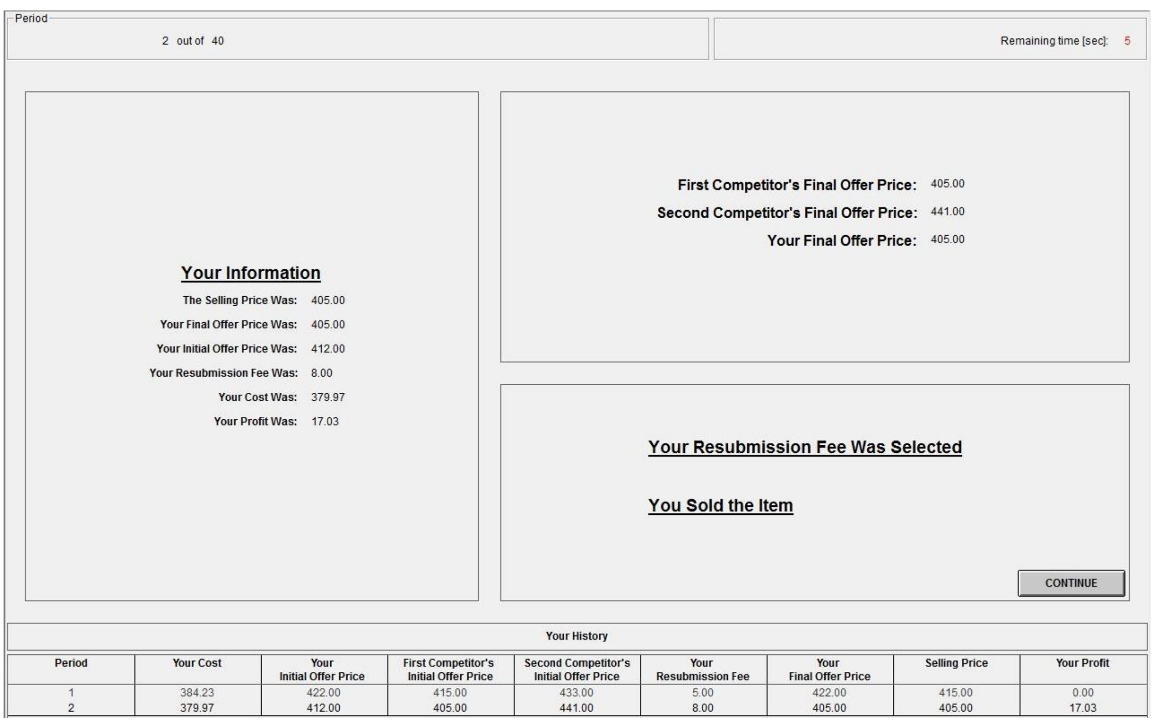
The auction ends when all three sellers have submitted their resubmission fees.

Again, only a seller whose resubmission fee is selected actually pays to match the lowest initial offer price, and this seller always sells the item. For all other sellers (including those who submit a resubmission fee of zero), their initial offer price *is* their final offer price. They might sell the item only if all sellers submit a resubmission fee of zero.

Information You Will See at the End of Each Period

At the end of each period, you will see messages telling you whether your resubmission fee was selected and whether you sold the item. You will also see summary information, including the selling price and your profit for the period.

At the bottom of most screens, you will see information from all previous periods. Below is a sample screen.



Summary

1. Your cost to produce the item is drawn randomly each period and is equally likely to be any number (to two decimal places) between \$375.00 and \$400.00, including \$375.00 and \$400.00.

2. The computerized buyer does not accept initial offer prices above \$475.00.

3. After submitting your *initial* offer price and learning the initial offer prices of the other two sellers, you compete with these sellers for the ability to match the lowest initial offer price and sell the item. You do this by submitting a resubmission fee of up to \$20.00. If your resubmission fee is above zero and the highest (or randomly chosen in the case of a tie), you are selected to match the lowest initial offer price and sell the item. If you are not selected to resubmit, your initial offer price *is* your final offer price.

4. If you submit a resubmission fee of zero, you will not be selected to resubmit. You might sell the item *only* if your initial offer price is lowest (or tied for lowest) *and* both your competitors also submit a resubmission fee of zero.

Example: Suppose your initial offer price is \$430.25, and the other two sellers' initial offer prices are \$420.50 and \$410.00. Then suppose that in Stage 2 your resubmission fee is \$18.00, and the other two sellers' resubmission fees are \$5.50 and \$13.40. Then your resubmission fee is selected, and your final offer price becomes \$410.00. You sell the item and receive a payment (the selling price) equal to \$410.00. Your profit is \$410.00 minus your cost, minus \$18.00.

Ready to Begin

If you have any questions, please raise your hand.

You will now go through a series of comprehension questions to make sure you understand the experiment. Again, raise your hand if you have any questions.

After everyone has correctly answered these questions, there will be three practice periods before the actual experiment starts. Profits from the practice periods do not count toward your total profit.

3 Chapter 2

Measuring Chinese Corruption With Household Data: A Stochastic Frontier Approach

Abstract

I present a novel method to detect corruption using only household data. I apply stochastic frontier (SF) analysis to measure the degree to which Chinese households with opportunities for corruption underreport their income in comparison with other households, assuming the resultant differential is illegal income. I compare my results and method to those of Zhong (2018), who uses the same data but another method to find that the true incomes of households with corruption opportunities are, on average, about 15% higher than their reported incomes, an estimate that Zhong cannot statistically test. SF analysis produces an estimate of about 10% that is statistically significant. Given similarities in our robustness checks, my results vindicate Zhong's approach, though only SF analysis 1) provides evidence of statistical significance, and 2) addresses endogeneity, which may explain the estimate difference. My method provides a cheap and easy way to quantify the relative corruption between groups, regions, and countries.

3.1 Introduction

Corruption scholars are always in search of objective measures of bribery and graft, often spending a good amount of time, energy, and money. In this paper I present a relatively cheap method to detect corruption using only household data. I apply stochastic frontier (SF) analysis to measure the degree to which Chinese households with opportunities to engage in corruption underreport their income in comparison with other households. I then assume the resultant differential is illegal income.

To study the same question Zhong (2018) borrows from Pissarides and Weber's (1989) method to measure tax evasion.[1] Using a Chinese household survey, Zhong first uses OLS to estimate an expenditure function for households without opportunities for corruption, then inverts the function to forecast the true income of households with opportunities. The difference between true income and reported income, he assumes, is the value of illegal income. Using an expenditure function for household equipment and services — the focus of this study — Zhong finds that Chinese households with corruption opportunities underreport their incomes by 14.6%.

Using the same dataset, I employ SF analysis, which is a statistical way to estimate a production frontier and each producer's efficiency based on that frontier.[2] I estimate a "production" frontier in which the output is household expenditure on equipment and services, and the input is household income, controlling for location and household characteristics. I then calculate the degree to which each household "overuses" income in its expenditure. I assume the differential in overuse between household types, those with and without opportunities for corruption, is a measurement of illegal income. I find that Chinese households with opportunities

[1] Several more-recent papers also inspired Zhong. See his introduction for a summary of the new strand of literature on forensic economics based on micro-level data.
[2] Stochastic frontier analysis is also commonly used to estimate cost and profit frontiers.

for corruption underreport their incomes by 10.3%, lower than Zhong's estimate but directionally consistent. Furthermore, when I conduct Zhong's robustness checks using the SF method, I detect mostly similar patterns.

The SF method makes two primary assumptions.[3] First, it assumes all households accurately report expenditures on some items, and that bribes in kind or other forms of corruption do not influence these expenditures, which are related solely to the household's true disposable income. Second, some household groups are not corrupt. Alternatively, the method could assume all groups corruptible if the goal is to measure how corrupt one group is compared to another. With these assumptions, I can estimate a reliable expenditure function and uncover illegal income.

The SF method has two substantial advantages over Zhong (2018). First, it can test the statistical significance of its measure of corruption, whereas Zhong cannot test the significance of his key parameters, and thus cannot test whether the amount of underreporting is significant.[4] Second, it partially corrects for endogeneity, whereas Zhong does not. On the other hand, a drawback of the SF method is that it employs a complicated maximum-likelihood (ML) estimation instead of OLS. As long as the ML function converges, however, both methods provide a cheap and fast way to measure corruption in comparison with other labor-intensive strategies. Most existing micro-level empirical analyses of corruption rely on administrative records, special-purpose surveys, or field experiments, which can be difficult or very costly to obtain. Most corruption studies in general rely on perception-based country-level corruption indices, which use corruption ratings based on expert opinions or surveys of business executives. Several studies, however, have found that perceptions of corruption are not perfect measures of

[3] Zhong makes the same assumptions.
[4] Zhong acknowledges this in his paper. In his footnote 17 he says one possible way of estimating standard errors might be to design an appropriate bootstrapping strategy.

actual corruption (e.g., Olken 2009 and Donchev and Ujhelyi 2014). This paper adds to this literature by exploiting a corruption-related question in the same household survey. Using both Zhong's method and the SF method to measure corruption, I find that households that report being less concerned about corruption live in areas with much more corruption. This may be the case because the least informed citizens live in the most corrupt areas — in which officials and the press do little to expose that corruption to the community.

Gorodnichenko and Peter (2007) were the first to measure corruption using only household data. They found that public employees in the Ukraine were underpaid compared with their private-sector counterparts. The wage gap remained even after they controlled for observable and unobservable characteristics, corrected for endogeneity, and accounted for differences in working hours, satisfaction, fringe benefits, job security, bonuses, and secondary employment. At the same time, the levels of expenditures and asset holdings reported by officials and their private-sector counterparts were the same, indicating bribery as long as the labor market was free to reach an equilibrium in which total worker compensation was equalized across sectors. Gorodnichenko and Peter provided evidence that no mobility constraints existed. Because Chinese public employees are paid more, on average, than their private counterparts, I cannot apply Gorodnichenko and Peter's method.

To the best of my knowledge, only two other papers measure corruption using household data. Both rely on methods different from those used in this paper. Saha et al. (2014) analyze the private-public wage gap in India to impute the existence of the unreported income. Because the gap is both positive and negative, depending on income level, the authors use a quantile regression technique. Their hypothesis (for which they find support) is that if they observe positive gaps for both wages and nondurable-good expenditures, but no such gap for durable-good expenditures, then they provide evidence of unreported earnings for official households. Nguyen (2017) examines whether natural disasters in Vietnam have the same effect on the income differential between official and nonofficial households as it does on the expenditure differential. Relying on the exogeneous effects of disasters on income and expenditure, Nguyen runs a series of regressions with panel data and fixed effects. Included in his analysis is a regression for disaster-affected areas only. The dependent variable is the ratio of expenditure to income, logged to represent the expenditure-income gap. Among the independent variables is an indicator of whether the household contains an official. The coefficient of this variable is positive and significant, implying that the official households in the affected areas spend more, even after controlling for reported income. Nguyen assumes this extra expenditure is funded by illegal income.

Also to the best of my knowledge, few micro-level papers have employed SF analysis to uncover the underreporting of income. A noteworthy example is Shonkwiler et al. (2011). They use SF analysis to measure the underreporting of remittances in Armenia. After controlling for this, they find a strong negative impact of remittances on incentives to work.

3.2 Method

3.2.1 Overview

As discussed above, I estimate a production function in which the output is household expenditure on equipment and services, and the input is household income, with controls for location and household characteristics. I then calculate the degree to which firms "overuse" income. I hypothesize that, in comparison with households without opportunities for corruption, households with opportunities seem to require less *reported* income to buy the same amount of goods and services (i.e., they are more "efficient") — but only because they do not report their illegal income. That is, I assume the differential in overuse between the two household types is a measurement of the illegal income obtained through corruption such as bribery and graft.

The SF approach requires that households accurately report expenditure for at least one item. Disposable income must determine this expenditure, meaning bribes in kind must not include this item. The raw data that I use — the same as Zhong's — are the 2002 Chinese Household Income Project (CHIP) data, which include household expenditures on a wide variety of items. Zhong claims that people in China commonly bribe with food, meals, alcohol, and cigarettes. Employer-provided fringe benefits — not just disposable income — partly determine household spending on healthcare, transportation, communication, and education. Zhong concludes that true disposable income most heavily influences expenditures on clothing and on household equipment and services. Favor-seekers hardly ever use clothing as bribes because of the difficulty in obtaining accurate information about the recipient's size and tastes. Furthermore, households are unlikely to claim clothing expenditure as a tax-deductible business expense, and employers are unlikely to subsidize it. Regarding expenditure on household equipment and services, a major component of this budget item (spending on household services like babysitting and house cleaning) is difficult to use as bribes, and people in China are much less inclined to bribe with small items of household equipment. Reports issued by the United Nations Office on Drugs and Crime (UNODC) empirically support these assumptions.[5]

[5] Zhong writes: "UNODC supported seven western Balkans countries or areas (Albania, Bosnia and Herzegovina, Croatia, Kosovo, Montenegro, Serbia, and Macedonia) to conduct large-scale surveys on corruption in 2010. A latterly published report indicates that bribery is prevalent in this region, and bribes are paid in the form of cash (66%), food and drink (22%), valuables (5%), exchange of services (3%), and other goods (4%) (UNODC 2011a). In Bosnia and Serbia, shares of bribes paid in the form of 'other goods' are only 2% (UNODC 2011b,c). This implies that clothing and household equipment are rarely used as in-kind bribes."

SF analysis estimates production inefficiency.[6] There are two possible measures of inefficiency: output-oriented (OO) and input-oriented (IO). The IO measure starts from the fact that if a producer is not efficient, it does not use inputs effectively. That is, there are slacks in the inputs, and producers can reduce input usage without reducing output. Consequently, the IO measure is practical and intuitive when output is exogenous and the objective is to maximize the proportional reduction in input usage without changing output. This potential maximum reduction — measured as a percentage of all inputs — is the IO measure. By contrast, OO technical inefficiency measures the potential increase in output without increasing the input quantities. Alternatively, it can be viewed as a measure of output loss resulting from failure to produce the maximum possible output permitted by the technology. Thus, the OO measure is intuitive when the inputs are exogenously given, and the objective is to produce as much output as possible.

To summarize, a production plan is technically inefficient if a higher level of output is attainable for the given inputs (OO measure), or the observed output level can be produced using fewer inputs (IO measure). In my context I assume household expenditure is a response to household income, so I choose the OO measure of inefficiency, assuming income is exogenous.[7] Thus expenditure is the dependent variable, and income is the independent variable. But I do not wish to measure the magnitude by which households could increase their expenditures given their income (the OO measure). I wish to measure each household's overuse of income given its expenditure choice. This second type of IO measure is subtly different from the one already

[6] Kumbhakar et al. (2015) provide a good description of SF analysis.

[7] For his main analysis Zhong also assumes income to be exogenous, though he tries to correct for endogeneity by using three dummies as instruments in a two-stage least square (2SLS) estimation. Hausman tests indicate that the 2SLS results are indistinguishable from the more efficient OLS results. In general, households probably often do earn more so they can spend more.

described, which would require income to be the dependent variable. That is, this second type of IO measure is defined as the first type but estimated differently, leading to a different numerical value.[8] Fortunately, as I explain below, I can easily derive this second type of IO measure from the OO measure.

Estimates of inefficiency are conditional on the given technology (production frontier). An input-output combination may appear inefficient for one technology, but it could be efficient with respect to a different technology. The implication for empirical analysis is that when estimating the technical inefficiencies of different producers, it is important that they are estimated with respect to the appropriate technology. In my application, technology is expenditure preferences: both types of households must have similar preferences for equipment and services. In the next section, I test for equal preferences and partly control for possible differences through propensity scoring.

3.2.2 Functional Form

To model the production function, I use the Cobb-Douglas functional form:

$$y = f(x) = A \prod_{j=1}^{J} x_j^{\beta_j}$$

$$\Rightarrow \quad \ln y = \beta_0 + \sum_j \beta_j \ln x_j \,, \tag{1}$$

where $\beta_0 = A$.

In my application, income is the only input, so the functional forms simplify to

[8] Kumbhakar et al. (2015) note that this issue is like regressing y on x and x on y, in which case the estimated slope coefficients are not reciprocals.

$$y = f(x) = Ax_1^{\beta_1}$$

$$\Rightarrow \quad \ln y = \beta_0 + \beta_1 \ln x\,. \tag{2}$$

OO technical inefficiency enters the production function as the u term in the following expression:

$$y = f(x)e^{-u}. \tag{3}$$

The logarithm of the production function makes $-u$ an additive term to the corresponding neoclassical specification

$$\ln y = \ln f(x) - u. \tag{4}$$

With IO technical inefficiency (η) in the production function, the expression is

$$y = f(xe^{-\eta}), \tag{5}$$

which, in logarithmic form, generally becomes

$$\ln y = \beta_0 + \sum \beta_j \ln x_j - \left(\sum_j \beta_j\right)\eta, \tag{6}$$

which is essentially the same as the OO model with the reparameterization

$$u = \eta \sum\nolimits_j \beta_j\,. \tag{7}$$

Thus, once I estimate u, I can easily obtain η from the relationship expressed in (7).[9] In my application this means dividing u by the coefficient of log income. In general terms, producers

[9] The intuition is as follows: think of a single y and single x. Because the relationship between $\ln y$ and $\ln x$ is linear, the vertical distance from a point below the line (which measures OO inefficiency, u) is the product of the horizontal distance (IO inefficiency, η) times the slope (β), that is, $u = \eta\beta$.

on average use $(\eta \times 100)\%$ more input than necessary due to technical inefficiency. I hypothesize that households with corruption opportunities overuse income to a lesser degree only because they do not report their illegal income. So my method for measuring illegal income consists of estimating u_i for each household i (using the method described below), deriving an observation-specific η_i from u_i, then calculating the average η_i for both types of households. The difference between these averages is my measurement of illegal income as a percentage of reported income.

As already mentioned, I assume that income is exogenous in the expenditure function, even though it is endogenous. Nevertheless, the SF method addresses this endogeneity by construction: the corruption measurement is the *difference* in two measurements of overuse (the two η_i averages), each constructed using the same (likely biased) coefficient of log income. If the two overuse measures are biased similarly due to this biased coefficient, the corruption measure should be mostly unbiased due to income endogeneity (though, as I discuss later, I must correct for possible bias in u_i itself due to heteroskedasticity). In the results section I report the overuse of income for each household type, but due to their biases, I analyze only their differences.

I choose to study household expenditure on equipment and services because the proper specification for such expenditure, the Working-Leser linear (piglog) formulation, mirrors the Cobb-Douglas form with its assumption of constant elasticity. Zhong also studies expenditures on equipment and services, but his most prominent analysis is based on clothing expenditure. Citing several empirical studies, Zhong reports growing evidence that clothing expenditure functions require a quadratic term for log income. Given this, I could have used the translog functional form for the production function. Estimation of the overuse measurement (η_i), however, would have been more complicated. To simplify the analysis without loss of generality, I choose to focus on equipment and services, the data for which support the omission of a quadratic term.[10]

3.2.3 Estimation of Mean Inefficiency

To estimate OO technical efficiency, I can specify a stochastic production frontier model with OO technical efficiency as

$$\ln y_i = \ln y_i^* - u_i, \qquad u_i \geq 0, \tag{8}$$

$$\ln y_i^* = f(\boldsymbol{x}_i; \boldsymbol{\beta}) + v_i, \tag{9}$$

where the subscript *i* denotes observations (households), y_i is a scalar of observed output (expenditure), $\boldsymbol{x}_i$ is a J×1 vector of the one input variable (income) and the controls, $\boldsymbol{\beta}$ is a vector of the corresponding coefficient vector, v_i is a zero-mean random error, and $u_i \geq 0$ is production inefficiency. Equation (9) defines the stochastic production frontier function. Given $\boldsymbol{x}$, the frontier gives the maximum possible level of output, and it is stochastic because of v_i. Given that $u_i \geq 0$, observed output (y_i) is bounded below the frontier output level (y_i^*). I also can write the model in the form:

$$\ln y_i = f(\boldsymbol{x}_i; \boldsymbol{\beta}) + \epsilon_i, \tag{10}$$

$$\epsilon_i = v_i - u_i, \tag{11}$$

where ϵ_i is the error term often called the composed error term.

[10] If I ignore the misspecification and use the same method described in this paper with clothing expenditure, I find that Chinese households with corruption opportunities underreport their income by 17.3% at the 1% level of significance. Using the correct specification, Zhong finds 18.9% with unknown significance. In Appendix A I fail to replicate his results based on clothing expenditure.

I construct a statistic to estimate u_i as follows. First, recall that I assume the Cobb-Douglas functional form for the production frontier $f(x)$. The estimation of the model then involves (i) estimating the parameters of the frontier function $f(x)$, and (ii) estimating inefficiency. To estimate $f(x)$ I first impose distributional assumptions on the error components, then derive the log-likelihood function of the model and use numerical maximization procedures to obtain the ML estimates of the model parameters. A zero-mean normal distribution for v_i is widely accepted in this context (Kumbhaker et al. 2015). The choice of distributional assumption for the random variable u_i is more the issue at stake. The distribution must be in the nonnegative domain, and its joint distribution with v_i would ideally have a closed form. The literature has identified several such distributions. In the first estimation of the SF model with distributional assumptions on v_i and u_i, Aigner et al. (1977) assumed a half-normal distribution for u_i. The half-normal distribution has a single parameter and is thus relatively easy to estimate. Subsequent developments in the literature have suggested more flexible (but harder to estimate) distribution functions such as the truncated-normal distribution with or without scaling properties (Stevenson 1980; Wang and Schmidt 2002). For my application, however, a half-normal distribution is appropriate.[11] Recall that I estimate u_i and then divide by a scalar to obtain η_i, which measures how households overuse income. Households seem to overuse income because other households do not report their full income and thereby seem "efficient" to varying degrees. In this way the most dishonest households define the frontier. Studies have shown that survey respondents tend to underreport income (Ravallion 2003; Freund and Spatafora 2008). I therefore expect most households to be efficient, which would result in a clustering of η_i near a

[11] In his method Zhong also relies on a distributional assumption. He assumes a log-normal distribution truncated at 1 for his corruption parameter, which measures the percentage by which public officials' reported income must be inflated to equal their true income. One drawback of his assumption is that no public official can earn only honest income (see his footnote 6).

calibrated zero, with a tail to the right representing the households that report honestly to varying degrees.[12]

Based on (8) and (9), a production SF with a normal distribution on v_i and a half-normal distribution on u_i is represented as the following:

$$\ln y_i = \ln y_i^* - u_i, \tag{12}$$

$$\ln y_i^* = \boldsymbol{x_i \beta} + v_i, \tag{13}$$

$$u_i \sim i.i.d. N^+(0, \sigma_u^2), \tag{14}$$

$$v_i \sim i.i.d. N(0, \sigma_v^2), \tag{15}$$

where v_i and u_i are distributed independently of each other. The $\boldsymbol{\beta}$, σ_u^2, and σ_v^2 are the parameters to be estimated. Equation (14) assumes that the inefficiency effect follows a half-normal distribution. One way to derive the half-normal distribution is to treat it as the nonnegative truncation of a zero-mean normal distribution. I shall denote the distribution derived in this way as $N^+(0, \sigma_u^2)$, where σ_u^2 is the variance of the normal distribution before truncation. Suppose that a random variable Z has a normal distribution $z \sim N(\mu, \sigma_z^2)$ with the probability density function denoted by $g(z)$. If it is truncated from above at the point α so that $z \geq \alpha$, then the density function of z, $f(z)$, is

$$f(z) = \frac{g(z)}{1 - \Phi\left(\frac{\alpha - \mu}{\sigma_z}\right)} = \frac{\frac{1}{\sigma_z}\phi\left(\frac{z - \mu}{\sigma_z}\right)}{1 - \Phi\left(\frac{\alpha - \mu}{\sigma_z}\right)}, \qquad z \geq \alpha, \tag{16}$$

[12] According to Zhong, there is no formal estimate of the size of China's shadow economy. A large one would further exacerbate underreporting.

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the probability density and probability distribution functions, respectively, for the standard normal variable.[13] The density function of u_i in (14) can then be obtained by setting $\mu = 0$ and $\alpha = 0$ in the above equation to give the following:

$$f(u_i) = \frac{\frac{1}{\sigma}\phi\left(\frac{u_i}{\sigma}\right)}{1 - \Phi(0)} = \frac{2}{\sigma}\phi\left(\frac{u_i}{\sigma}\right) = 2(2\pi\sigma^2)^{-\frac{1}{2}}\exp\left(-\frac{u_i^2}{2\sigma^2}\right), \qquad u_i \geq 0. \tag{17}$$

The log-likelihood function based on (12)-(15) for each observation i is

$$L_i = -\ln\left(\frac{1}{2}\right) - \frac{1}{2}\ln(\sigma_v^2 + \sigma_u^2) + \ln\phi\left(\frac{\epsilon_i}{\sqrt{\sigma_v^2 + \sigma_u^2}}\right) + \ln\Phi\left(\frac{\mu_{*i}}{\sigma_*}\right), \tag{18}$$

where

$$\mu_{*i} = \frac{-\sigma_u^2\epsilon_i}{\sigma_v^2 + \sigma_u^2}, \tag{19}$$

$$\sigma_*^2 = \frac{\sigma_v^2\sigma_u^2}{\sigma_v^2 + \sigma_u^2}. \tag{20}$$

For detailed derivations, see Appendix A of Kumbhakar et al. (2015). The log-likelihood function is then the observational sum of (18), which can then be numerically maximized to obtain the estimates of the model parameters. There is, however, a computational problem. The variance parameters, σ_u^2 and σ_v^2, must be positive, but an unconstrained numerical maximization would not guarantee positive estimates. To ensure that the variance parameter estimates are positive, researchers in the early literature often used the following parameterization scheme for the unconstrained numerical maximization:

[13] See Johnson et al. (1995).

$$\sigma_u^2 = \exp(w_u), \tag{21}$$

$$\sigma_v^2 = \exp(w_v), \tag{22}$$

where w_u and w_v are unrestricted constant parameters (Kumbhakar et al. 2015). As I explain momentarily, I choose to go step further and correct of heteroskedasticity.

3.2.4 Correcting for Heteroskedasticity

The original half-normal model of Aigner et al. (1977) assumes that the v_i and the pretruncated u_i are homoskedastic, that is, both σ_v^2 and σ_u^2 are constants. Caudill and Ford (1993), Caudill, Ford, and Gropper (1995), and Hadri (1999) consider models in which these random variables are heteroskedastic. Unlike a classical linear model in which heteroskedasticity affects only the efficiency of the estimators and not their consistency, ignoring heteroskedasticity in the SF framework leads to inconsistent estimates (Wang and Schmidt 2002). Kumbhakar and Lovell (2000, Section 3.4) provide a detailed discussion on the consequences of ignoring the heteroskedasticity, assuming v_i and u_i are heteroskedastic. Ignoring the heteroskedasticity of v_i still gives consistent estimates of the frontier function parameters ($\boldsymbol{\beta}$) except for the intercept, which is downward-biased. Estimates of the technical efficiency are biased. Ignoring the heteroskedasticity of u_i causes biased estimates of the frontier function's parameters as well as the estimates of technical efficiency. Caudill and Ford (1993), Caudill, Ford, and Gropper (1995), and Hadri (1999) propose the heteroskedasticity can be parameterized by a vector of observable variables and associated parameters. For instance, $\sigma_{u,i}^2 = \exp\left(\boldsymbol{z}_{u,i}; \boldsymbol{w}_u\right)$, where $\boldsymbol{z}_{u,i}$ is an $m \times 1$ vector of variables including a constant of 1, and $\boldsymbol{w}_u$ is the $m \times 1$ corresponding parameter vector. The exponential function is used to ensure a positive estimate of the variance parameter. Therefore, the parameterizations are

$$\sigma_{u,i}^2 = \exp(\boldsymbol{z'}_{u,i}\boldsymbol{w}_u)\,, \tag{23}$$

$$\sigma_{v,i}^2 = \exp(\boldsymbol{z'}_{v,i}\boldsymbol{w}_v)\,. \tag{24}$$

The vectors $\boldsymbol{z}_{u,i}$ and $\boldsymbol{z}_{v,i}$ may or may not be the same vector, and they may contain all or part of the $\boldsymbol{x}_i$ vector. In my application I am interested in the difference between the mean u_i (divided by the coefficient of log income) for households with opportunities for corruption and the mean u_i (divided by the same scalar) for all other households. My main concern, therefore, is how the bias in u_i due to heteroskedasticity differs between household types, so I parameterize both $\sigma_{u,i}^2$ and $\sigma_{v,i}^2$ with an indicator variable for households with opportunities for corruption.

The log-likelihood function of the heteroskedastic model is the same as in (18), except that I now use (23) and (24) instead of (21) and (22) in place of σ_u^2 and σ_v^2, respectively, in the log-likelihood function. All the parameters of the model are estimated at the same time via the ML method.

3.2.5 Estimation of Observation-Specific u_i

After I estimate the model parameters, I can estimate the observation-specific u_i. Although the definition of this index is intuitive, estimating the index for each observation is less straightforward. To see this, note that $u_i \sim N^+(0, \sigma_u^2)$. The ML estimation of the model yields the estimated value of σ_u^2, which provides information about the shape of the half-normal distribution of u_i. This information is all I need to find the average technical inefficiency of the sample. However, I am interested in the u_i of each observation, so this information on σ_u^2 is not enough because it does not contain any household-specific information. The solution, first proposed by Jondrow et al. (1982), is to estimate u_i from the expected value of u_i conditional on the composed error of the model, $\epsilon_i \equiv v_i - u_i$. This conditional mean of u_i given ϵ_i gives a point estimate of u_i. The composed error contains household-specific information, and so the conditional expectation yields the observation-specific value of the inefficiency.[14] Jondrow et al. (1982) show that the density function of $(u_i|\epsilon_i)$ is $N^+(u_{*i}, \sigma_*^2)$, based on which, the equation of $E(u_i|\epsilon_i)$ is (see Appendix B of Kumbhakar et al. 2015):

$$E(u_i|\epsilon_i) = \frac{\sigma_* \phi\left(\frac{\mu_{*i}}{\sigma_*}\right)}{\Phi\left(\frac{\mu_{*i}}{\sigma_*}\right)} + \mu_{*i}, \tag{25}$$

where μ_{*i} and σ_* are defined in (19) and (20). I substitute into the equation ML estimates of the parameters to obtain the empirical estimate of inefficiency, which is guaranteed to be nonnegative. Estimates of observation-specific inefficiency assume the model parameters are known and given, while in fact I estimate them with uncertainty.

3.3 Data and Variables

I use the same raw dataset as Zhong: the 2002 CHIP, which was conducted in 2003 for the prior year. Ideally, I would use his sample from the raw data, but his dataset is unavailable.[15] The 2002 CHIP is a rich dataset collected primarily by the Chinese National Bureau of Statistics. I use only the urban survey, which covers 6,835 households and 20,632 individuals in 12 provinces.[16] Zhong claims he selected his sample according to three criteria: (i) the household contains two adults; (ii) the second adult is the wife; and (iii) the head of household is employed (excluding

[14] Kumbhakar et al. (2015) likens this to extracting signal from noise.

[15] In personal correspondence, Zhong informed me that he has only a one-page text file with a few hints. The text file explains how he divided his sample into corrupt and noncorrupt households (and further subsamples), but I was left to my own devices in constructing the variable for disposable income. I did my best to recreate his sample using this text file and the paper itself. I believe I was mostly successful due to the similarity of our summary statistics, as shown in Tables AA2-AA3 in Appendix A. I am also mostly successful in replicating Zhong’s results based on expenditures on equipment and services. However, I could not replicate Zhong’s results based on clothing expenditure, an issue I explore in Appendix A.

[16] These provinces are Anhui, Beijing, Chongqing, Gansu, Guangdong, Henan, Hubei, Jiangsu, Liaoning, Shanxi, Sichuan, and Yunnan.

self-employed and retired people).[17] The purpose of this sample selection process, Zhong says, is to avoid differences owing to the preferences of various other household groups, such as single parents, the self-employed, and pensioners. After the selection process, Zhong is left with 4,213 households. When I attempt to the selection criteria, I encounter a problem: when I drop from the raw CHIP dataset only households not containing exactly two adults, I am left with only 3,866 households, far below Zhong's count even before I implement the other criteria. I therefore allow other adults to reside in the household while following Zhong's other criteria. My final sample contains 4,395 households.

I calculate illegal income by estimating how households with opportunities for corruption differ in their underreporting of income when compared with other households. If the two types of households differ in their underreporting for another reason — tax evasion — then my results are confounded. I address this concern as Zhong does. I exclude all self-employed respondents from my sample. For wage earners, it is relatively difficult to underreport employment income because of employer reporting. Moreover, unlike most developed countries, the personal income tax is not a major source of government revenue in China.[18] I therefore believe that tax-evasion motives affect my results only negligibly, although this cannot be tested with the CHIP data.

Zhong divides China's state sector into three segments: government, budgetary organizations (BOs), and state-owned enterprises (SOEs). According to him (without citations), BOs are like nonprofit and nongovernment organizations but are financed by and fully accountable to the

[17] Zhong's criteria suggest that his sample includes only households with exactly two adults, and footnote 10 suggests likewise. Zhong's note for his Table 1 defines adults as people age 18 and older.
[18] According to Zhong, receipts from the personal-income tax accounted for just 6.9% of China's total revenue in 2002 (China State Statistical Bureau 2003). Zhong made the following calculations using the 2002 Chinese personal-income-tax schedule: for those respondents who reported incomes in the 2002 CHIP data, only 41.3% were required to pay personal income tax, and almost all of them (39.8%) faced a marginal tax rate below 10%.

government. BOs mainly engage in the sectors of education, scientific research, cultural activities, and health services. Examples include primary and middle schools, universities, research institutes, news agencies, healthcare institutions, television networks, publishing houses, museums, parks, and state-controlled associations. Ostensibly, SOEs are profit-driven institutions that are self-financing, but some of them are highly subsidized. They are managed by government-appointed directors, and those directors often switch periodically between being government officials and firm managers. The government grants significant freedom to directors and managers of BOs and SOEs, which have more relaxed financial constraints compared to private firms. The mass media in China frequently reports on these officials' corrupt behavior, such as using public money in exchange for personal benefits.

One question in the 2002 CHIP asked respondents about the type of organization for which they worked. The answer options included: government, budgetary organization, private or public enterprise, and other. There was also a question on the occupational rank of the respondent, for which the answer options included: manager of a private firm, self-employed, professional, director of government agency/institution/enterprise, departmental director of government agency/institution/enterprise, clerical/office staff, skilled worker, unskilled worker, sales clerk or service worker, and other.

In what follows, my analysis is an emulation of Zhong's unless I note otherwise. I start with the assumption that no household members who are employed in the private sector can engage in corruption (hereafter called the noncorrupt group), but that all government employees and staff at the management level in BOs and SOEs can do so (hereafter called the corrupt group). According to Zhong, most junior staff in BOs and SOEs do not provide services directly to the population and have limited discretion in allocating public money. Hence I assume that they do not have the opportunity to engage in corruption. These assumptions may lead to an underestimation of the extent of corruption if private-sector employees or junior-level staff in BOs and SOEs engage in corrupt activities. In this case, the estimates of corruption serve as a lower bound. Alternatively, I can consider my corruption estimate as a measure relative to the reference group. In the results section I consider how changes in defining the membership of the noncorrupt group affect the results and try to find a “cleaner” reference group. I define households with one or more members in the corrupt group as households with opportunities for corruption, and all other households as households without them. I (Zhong) classify 958 (879) observations as households with opportunities for corruption, and 3,437 (3,334) observations as households without opportunities for corruption.

3.3.1 Variables

For comparability, I use the same variables as Zhong. Dependent variables include the logarithm of household expenditure on household equipment and services (my main dependent variable), the logarithm of household expenditure on clothes (to replicate Zhong’s result in the appendix), and the logarithm of household expenditure on food (for comparison’s sake). The explanatory variable is the logarithm of household income, which is the sum of a household’s cash income and other incomes in kind from all sources, excluding taxes and fees.[19] All expenditure and income variables are measured in Chinese Yuan (RMB). Control variables linked to assets include the number of cars, motorcycles, telephones, washing machines, televisions, refrigerators, and computers owned by the household. Demographic control variables include the

[19] Pissarides and Weber (1989) use a measure of permanent income in the expenditure equation to address income uncertainty. Zhong and I choose instead to use current income to simplify our analysis. Citing several papers, Zhong notes that whether households smooth their consumption due to predictable changes in income is a contentious question. Stephens (2008) shows that household nondurable consumption is sensitive to changes in income in the presence of credit-market imperfections, which are widespread in developing countries (see Zhong for citations).

age of the head of household, the square of the age of the head of household, the number of children in the household, an indicator whether the head of household was in poor health, and years of education attained by the head of household and the spouse. Control variables concerning household characteristics include indicators for whether the head of household worked in a BO, the spouse worked in a BO, the household had a mortgage, the household needed to pay the cost of education for the children, the household needed to pay rent for accommodation, the head of household was a member of the Communist Party, the spouse was a member of the Communist Party, the firm at which the head of household was employed was not making a loss, the head of household was a director (or manager) of a government agency/institution/enterprise, and the spouse is working. Also included is a set of dummy variables for the province in which the household resided. Table 1 reports the statistics for the whole sample, households with corruption opportunities, and households without corruption opportunities. Table 2 reports the summary statistics for the two subsamples of the noncorrupt group: private-sector employees, and junior-level staff in BOs and SOEs, which I will use for a robustness check.[20]

The statistics for households with corruption opportunities and households without corruption opportunities indicate that their characteristics are different. In terms of income, consumptions, and household assets, the corrupt group is considerably better off than the noncorrupt group. My method relies on the relationship between income and expenditures, so the difference between the income distributions is a concern. Correct results clearly depend on getting Engel effects right, the more so as the income distributions of the two groups differ. I draw kernel density plots

[20] Tables AA2-AA3 in Appendix A contain these same summary statistics next to the corresponding figures reported by Zhong. All figures are similar except for the percentage of spouses who work. This indicates that I was mostly successful in recreating his sample from the raw data.

of these two income distributions in Figure 1, which shows that the income distribution for the corrupt group dominates that for the noncorrupt group.[21] However, the ranges of the two distributions are similar. In this case, if I can ensure that the two groups of households are similar in preferences, then the estimated income-expenditure relationship can be used to fully reflect the true level of corruption. Figure 2 presents kernel density plots of distributions of expenditures on household equipment and services for the two groups. The expenditure distribution for the corrupt group dominates that for the noncorrupt group.

3.4 Empirical Results

3.4.1 Accounting for Preference Heterogeneity

My analysis has not yet accounted for preference heterogeneity. If corruptible households, compared to other households, have different preferences for equipment and services, then my measurement of illegal income will capture both preference heterogeneity and corruption. For example, officials might enjoy more generous fringe benefits and added job security, and these things could lead to increased spending on non-necessities like house-cleaning services. Later I test this hypothesis by adding two variables to the frontier estimation that could indicate a difference in preferences. Still, there could be other reasons for heterogeneous preferences. For example, people with different levels of income, wealth, education, household characteristics, location of residence, and other socioeconomic characteristics may have different consumption preferences. If some corruptible households, compared to other households, are very different in those factors, then my corruption interpretation could be incorrect. To mitigate this problem, I use propensity scoring to restrict those factors in a common support. I calculate the propensity

[21] Kolmogorov-Smirnov tests support my comparisons at the 1% level regarding both income distributions and expenditure distributions of the two household types.

score for each household and then restrict the sample to areas of common support. I estimate a probit model in which the outcome variable is an indicator of whether the household is in the corrupt group. Independent variables include log of income and some other variables that may affect household consumption.[22] By excluding those observations that are not in the region of common support, I (Zhong) further drop(s) 95 (85) households.

3.4.2 Main Results

Before I estimate illegal income, I follow the SF literature by testing for the one-sided error term, as explained in Appendix B. The pre-estimation tests (Equations AB1-AB2) strongly support the existence of a one-sided error term, as does a post-estimation test (AB3).

Using the method outlined in the previous section I proceed to estimate the model of Equations (12)-(15) with parametric distributional assumptions on v_i and u_i, assuming normal and half-normal distributions, respectively, and correcting for heteroskedasticity. See Table 3 for the ML estimation. The coefficients of the indicator variable for households with corruption opportunities in both $\sigma^2_{u,i}$ and $\sigma^2_{v,i}$ are statistically significant, implying that the inclusion of this variable in the model is supported by the data (See Equations [23]-[24] and the related discussion). I test for heterogenous preferences between the household types by adding two variables to the frontier estimation: dummy interaction variables for noncorrupt households in which the head of household or the spouse was employed in a junior staff position in a BO. The government and BOs, Zhong claims, have very similar characteristics, including working hours,

[22] See Zhong for a more thorough explanation of our application of propensity-score matching. Independent variables in the probit model include log income, age of the household head, number of children, household head's years of education, spouse's years of education, need to pay a mortgage, need to pay education costs for children, need to pay rent for accommodation, household head is a member of the Communist Party, household head is in a BO, spouse is a member of the Communist Party, spouse is in a BO, spouse is working, and the number of the following assets owned by the household: cars, motorcycles, telephones, washing machines, televisions, fridges, and computers.

fringe benefits, labor-force composition, organizational norms, and morale. However, as shown in Table 3, the coefficients of the two variables are close to zero and statistically insignificant. Perhaps high-ranking officials have different preferences? I address this concern by adding to the estimation a dummy variable for households in which the head of household was employed as a manager or director in his organization. The variable's coefficient is not statistically significant.

Next, I use Equation 25 to generate the estimated values of $E(u_i|\epsilon_i)$ evaluated at $\hat{\epsilon}_i$.[23] I divide these values by the coefficient of log income to obtain the observation-specific η_i, averages of which I calculate for both types of households. Table 4 reports these averages, denoted $\bar{\eta}_i^{nc}$ and $\bar{\eta}_i^{c}$, for the noncorrupt and corrupt households, respectively, along with all figures discussed below.[24] The difference between these averages is the illegal income of households with opportunities for corruption, expressed as a percentage of reported income. This estimate is 10.3% (p-value = 0.0000, according to a two-sided t-test) with a 95% confidence interval of 7.4% to 13.2%. Recall that this illegal income is specifically from bribery and graft under the assumptions that corrupt and noncorrupt households hide income identically for other reasons such as tax evasion and black-market participation. Also recall that I can interpret this result as only a relative corruption measure. My replication of Zhong's finding for this same percentage is 15.5% with unknown significance, close to his 14.6% (different perhaps due to our using slightly different samples). The failure of Zhong's method to correct for endogeneity may explain the difference between 10.3% and 15.5%. To explain why, I first give a brief description of his

[23] To estimate the model, I first set up the likelihood function by using the *sfmodel* Stata command written by the authors of Kumbhakar et al. (2015). I then estimate the model by using Stata's *ml max* command. Finally, I generate the observation-specific $E(u_i|\epsilon_i)$ using the *sf_search* command, also provided by Kumbhakar et al.

[24] As discussed in the method section, one should interpret the η_i averages with care because they are potentially very biased. The overuse estimates of 0.86 and 0.76 may seem high; however, Pissarides and Weber (1989) find in Britain that on average true self-employment income is 1.55 times as high as reported self-employment income. Here, I study corruption (and other informal income) in China instead of tax evasion in Britain, but the similarity in magnitude order is reassuring.

method. First, he regresses expenditure on log income, the controls, and a dummy indicating whether a household is corrupt. He then divides the coefficient of the dummy by the coefficient of log income. To this number he adds a (halved) measurement that I do not explain because of its relative complexity and because its inclusion usually alters the corruption estimate by only 1% to 3%. Finally, he exponentiates this sum using *e* as the base, giving the percentage by which public officials' reported income must be inflated to equal their true income. If measurement-error bias predominates, causing a downward bias for the coefficient of log income, Zhong's method overstates corruption. Because my corruption measurement is a difference in two similarly biased estimates, I partially correct for this.

Zhong reports the results of his corruption estimation using food expenditures. He claims the regression results for the quadratic food expenditure equation indicate that a quadratic term for log income is not statistically significant and unnecessary. He then proceeds to show that the estimated level of corruption is only 2.2% based on food expenditure, lending support to his theory that food is often used as bribes in kind and thus cannot be used to estimate corruption.[25] My regression results suggest strongly that the quadratic term is necessary, so for reasons given in the method section, I do not explore this issue with SF analysis.

3.4.3 Estimations With Different Reference Groups

I have assumed that all private-sector employees and junior-level staff in BOs and SOEs do not

[25] Zhong appears to make a mistake in his analysis. In his Table 4 he treats a key parameter (δ_1'), the coefficient for a dummy indicating that a household is corrupt, as if it were statistically significant, even though the table indicates that it is not. When I redo the calculations assuming the parameter is zero, I find underreporting of 0.08%, which only strengthens his argument. It is possible that table contains a typo regarding significance. However, my own regression indicates a significant coefficient of log income, 0.505 (close to Zhong's 0.468), with no significance for the δ_1' parameter. Regarding the 0.08% finding above, it is not clear to me whether I should even proceed with Zhong's method after finding δ_1' insignificant.

engage in corrupt activity, a strong assumption.[26] To test the extent to which my results depend on the classification of the noncorrupt group, I divide the noncorrupt group into two subgroups: one consisting of junior-level staff in BOs and SOEs (the BOSOE group), and the other comprising private-sector employees only (private group). I then estimate the corruption level of each group separately. The estimation results based on the corrupt group and the private group are reported in the section of Table 4 labelled "Reference group (1)," while the results of the estimation based on the corrupt group and the BOSOE group are reported in the section labelled "Reference group (2)." Varying the composition of the noncorrupt group does not qualitatively affect the conclusion regarding the higher level of corruption of government officials in my results (or Zhong's). However, compared with the corruption levels estimated with the main reference group, the estimated corruption levels are lower when I use the private group. This may indicate that private-sector employees have more opportunities for corruption than junior-level staff in BOs and SOEs.[27]

I divide the 12 provinces of our sample into two groups: rich provinces and poor provinces.[28] I then estimate the level of corruption for each group. In his paper Zhong finds that rich

[26] According to Zhong, some employees in BOs provide services directly to the population, and hence may have opportunities to engage in corruption. In addition, the Chinese media often report on employees in the private sector or SOEs who embezzle funds from stockholders, take bribes from suppliers, and steal money from their employers.

[27] According to Zhong, who cites China State Statistical Bureau 2003, privately owned firms are more heavily concentrated in sectors such as wholesale and retail trading, hotels and catering services, financial intermediation, and real estate, while publicly owned firms are more concentrated in heavy industries such as mining and power generation, as well as other sectors such as education and scientific research. Therefore, compared with junior-level staff in BOs and factory workers in SOEs, private-sector employees may have more opportunities to engage in corrupt activity, such as asking for rebates from suppliers, taking bribes from clients, and stealing money from employers. Another possible explanation for this result, Zhong notes, is that private firms may have stronger incentives to help their employees to avoid taxes, as the effective wage paid in private firms is higher.

[28] Zhong and I rank the 12 provinces by average household income, calculated from the 2002 CHIP data. We define the top six provinces as rich provinces. These provinces are Guangdong, Beijing, Jiangsu, Chongqing, Yunnan and Liaoning. We define as poor the other six provinces (Hubei, Shanxi, Anhui, Henan, Sichuan and Gansu).

provinces tend to have higher levels of corruption, and he speculates why. I find no remarkable difference, whether I use the SF method or Zhong's method (Table 4).

The CHIP survey asks a corruption-related question, to be answered by the household head or main member. The question asks the respondent to rank the three most important problems of his or her city. There are nine options, one of which is corruption. When I drop from my sample people who do *not* rank corruption as the No. 1 or No. 2 problem (leaving 58.9% of the sample), average underreporting shoots up to 14.2% (using the SF method with an exponential assumption) and 22.4% (using Zhong's method).[29] This group is labelled "Less Worried" in Table 4. When I consider only the opposite group, labelled "Worried About Corruption," the SF method detects no underreporting, and Zhong's method reports a low percentage. So the households experiencing the least corruption are the most concerned. One can only speculate about the reasons, as researchers have done with similar findings (see citations in the introduction). A greater concern for corruption perhaps leads to less corruption, or maybe more corrupt areas contain more corrupt respondents, who may be less inclined to list corruption as a top problem.[30] Or the result could be a vagary of the data: when I consider only households that list corruption as the No. 1 problem (874 households, 173 of which are corrupt), the SF method again reports significant corruption of 7.8%. Regardless, my finding is further evidence that researchers should be wary of relying too heavily on attitudinal survey questions when assessing the extent of corruption.

I try to classify a "cleaner" benchmark group according to respondents' occupations, labelled "Reference Group (3)" in Table 4. This reference group consists of workers who should have

[29] Zhong in his paper does not address this survey question.
[30] This latter conjecture seems unlikely because the proportion of households listing corruption as a top concern is the same regardless of household type: about 40%.

fewer corruption opportunities: unskilled workers, skilled workers, and service workers in both the private sector and SOEs. But the estimated level of corruption does not support this conjecture. This last finding underscores the measurement's roughness — the trade-off for its affordability and applicability.

3.5 Conclusion

Zhong claims to provide the first objective estimate of the extent of corruption in China (in 2002). According to his preferred measure of corruption, based on clothing expenditure, the true incomes of households with more opportunities for corruption are, on average, about 20% higher than their reported income. This estimate, however, is based on a quadratic expenditure function requiring a method that, at least according to my analysis in Appendix A, seems to be highly sensitive to sample selection. On the other hand, I easily replicate Zhong's estimate based on expenditure of household equipment and services, which does not require a quadratic term. That estimate indicates about 15% instead of 20%, although Zhong cannot say whether the estimates are statistically significant. Using the same household survey, I apply SF analysis to produce a statistically testable estimate corresponding to Zhong's 15% measurement. The SF method indicates about 10%, which is significant at the 1% level. Regarding the difference in our point estimates, I argue that my method alone may partially correct for bias in a key parameter — and thus may be closer to the truth. Even so, this paper primarily vindicates Zhong's approach to measuring corruption because most of our robustness checks follow similar patterns.

Clearly, a clean benchmark group is hard to find. These estimations, therefore, are best thought of either as lower-bound estimations or as relative measures of corruption (my preference). The value of the methods is their ease and affordability in comparison with micro-level analysis based on administrative records, special-purpose surveys, or field experiments.

With just the variables found in many household surveys, researchers can explore specific issues such as the relative corruption between specific groups or regions, or the correlation between a community's concern for corruption and the extent of corruption. Here, for example, my estimates seem to indicate a negative correlation. Researchers also can use the SF method to measure tax evasion. Policymakers, when allocating funds for fighting corruption, could benefit from objective measures of corruption's intensity, perhaps by region.

Regarding future research, Zhong points out that many methods exist for estimating the size of the underground economy, all of which could be applied to answer this paper's questions, and all which have their own advantages and weaknesses. China, however, is a peculiar country to test such methods. The public sector's predominance makes finding a clean benchmark group especially difficult. More interesting, perhaps, would be to see how estimates of corruption differ by country. In Redwine (2018), for example, I apply the SF method to study corruption in Indonesia.

References

Aigner, Dennis, CA Knox Lovell, and Peter Schmidt. "Formulation and estimation of stochastic frontier production function models." *Journal of Econometrics* 6.1 (1977): 21-37.

Caudill, Steven B., and Jon M. Ford. "Biases in frontier estimation due to heteroscedasticity." *Economics Letters* 41.1 (1993): 17-20.

Caudill, Steven B., Jon M. Ford, and Daniel M. Gropper. "Frontier estimation and firm-specific inefficiency measures in the presence of heteroscedasticity." *Journal of Business & Economic Statistics* 13.1 (1995): 105-111.

China State Statistical Bureau. *Statistical Yearbook of China, 2002.* Beijing: China Statistical Press, 2003.

Coelli, Tim. "Estimators and hypothesis tests for a stochastic frontier function: A Monte Carlo analysis." *Journal of productivity analysis* 6.3 (1995): 247-268.

D'Agostino, Ralph, and Egon S. Pearson. “Tests for departure from normality. Empirical results for the distributions of b_2 and $\sqrt{b_1}$.” *Biometrika* 60.3 (1973): 613-622.

Donchev, Dilyan, and Gergely Ujhelyi. “What do corruption indices measure?” *Economics & Politics* 26.2 (2014): 309-331.

Freund, Caroline, and Nikola Spatafora. “Remittances, transaction costs, and informality.” *Journal of Development Economics* 86.2 (2008): 356-366.

Gorodnichenko, Yuriy, and Klara Sabirianova Peter. “Public sector pay and corruption: Measuring bribery from micro data.” *Journal of Public Economics* 91.5-6 (2007): 963-991.

Hadri, Kaddour. “Estimation of a doubly heteroscedastic stochastic frontier cost function.” *Journal of Business & Economic Statistics* 17.3 (1999): 359-363.

Johnson, Norman L., Samuel Kotz, and N. Balakrishnan. *Continuous univariate distributions, vol. 2.* John Wiley & Sons, 1995.

Jondrow, James, et al. “On the estimation of technical inefficiency in the stochastic frontier production function model.” *Journal of Econometrics* 19.2-3 (1982): 233-238.

Kodde, David A., and Franz C. Palm. “Wald criteria for jointly testing equality and inequality restrictions.” *Econometrica: Journal of the Econometric Society* (1986): 1243-1248.

Kumbhakar, Subal C., Hongren Wang, and Alan P. Horncastle. *A practitioner's guide to stochastic frontier analysis using Stata*. Cambridge University Press, 2015.

Kumbhakar, Subal C., and CA Knox Lovell. *Stochastic frontier analysis*. Cambridge University Press, 2000.

Meeusen, Wim, and Julien van Den Broeck. “Efficiency estimation from Cobb-Douglas production functions with composed error.” *International Economic Review* (1977): 435-444.

Olken, Benjamin A. “Corruption perceptions vs. corruption reality.” *Journal of Public Economics* 93.7-8 (2009): 950-964.

Nguyen, Quang. “Do natural disasters open a window of opportunity for corruption?” *The Journal of Development Studies* 53.1 (2017): 156-172.

Pissarides, Christopher A., and Guglielmo Weber. “An expenditure-based estimate of Britain's black economy.” *Journal of Public Economics* 39.1 (1989): 17-32.

Ravallion, Martin. “Measuring aggregate welfare in developing countries: How well do national accounts and surveys agree?” *Review of Economics and Statistics* 85.3 (2003): 645-652.

Redwine, Arlo. “Measuring Indonesian Corruption With Household Data: A Stochastic Frontier Approach” *Working Paper* (2018).

Saha, Sarani, Poulomi Roy, and Saibal Kar. “Public and private sector jobs, unreported income and consumption gap in India: Evidence from micro-data.” *The North American Journal of Economics and Finance* 29 (2014): 285-300.

Schmidt, Peter, and Tsai-Fen Lin. “Simple tests of alternative specifications in stochastic frontier models.” *Journal of Econometrics* 24.3 (1984): 349-361.

Shonkwiler, J. Scott, David A. Grigorian, and Tigran A. Melkonyan. “Controlling for the underreporting of remittances.” *Applied Economics* 43.30 (2011): 4817-4826.

Stephens Jr, Melvin. “The consumption response to predictable changes in discretionary income: Evidence from the repayment of vehicle loans.” *The Review of Economics and Statistics* 90.2 (2008): 241-252.

Stevenson, Rodney E. “Likelihood functions for generalized stochastic frontier estimation.” *Journal of Econometrics* 13.1 (1980): 57-66.

United Nations Office on Drugs and Crime (UNODC). *Corruption in the Western Balkans: Bribery as Experienced by the Population.* Vienna: United Nations Office on Drugs and Crime, 2011a.

—— *Corruption in Bosnia: Bribery as Experienced by the Population.* Vienna: United Nations Office on Drugs and Crime, 2011b.

—— *Corruption in Serbia: Bribery as Experienced by the Population.* Vienna: United Nations Office on Drugs and Crime, 2011c.

Wang, Hung-Jen, and Peter Schmidt. “One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels.” *Journal of Productivity Analysis* 18.2 (2002): 129-144.

Zhong, Hai. “Measuring corruption in China: An expenditure-based approach using household survey data.” *Economica* 85.338 (2018): 383-405.

Appendix

Tables and Figures

Table 1: Mean of Variables Used in the Analysis

Variable	Full Sample	Household with corruption opportunities	Household without corruption opportunities
Household income	25,824.15	30,678.77	24,471.01
Log of household income	10.01	10.21	9.95
Clothing expenditure	1,996.86	2,591.22	1,831.19
Log of clothing expenditure	7.27	7.60	7.18
Food expenditure	6,807.73	7,384.59	6,646.94
Log of food expenditure	8.69	8.78	8.67
Household equipment and services	1,226.66	1,578.66	1,128.55
Log of household equipment and services	6.34	6.61	6.26
No. of cars owned by household	0.01	0.01	0.01
No. of motorcycles owned by household	0.20	0.21	0.19
No. of telephones owned by household	3.63	3.93	3.55
No. of washing machines owned by household	0.96	0.99	0.95
No. of televisions own by household	1.27	1.32	1.26
No. of fridges owned by household	0.89	0.91	0.89
No. of computers owned by household	0.27	0.32	0.26
Age of household head	43.53	44.08	43.38
No. of children in household	0.61	0.58	0.62
Household head has poor health	0.04	0.04	0.04
Years of education received by household head	11.39	12.55	11.06
Years of education received by spouse	10.76	11.73	10.49
Need to pay mortgage	0.04	0.06	0.03
Need to pay education costs of children	0.14	0.16	0.13
Need to pay rent for accommodation	0.16	0.13	0.17
Household head is member of Communist Party	0.38	0.62	0.32
Spouse is member of Communist Party	0.22	0.41	0.17
Spouse is working	0.72	0.79	0.71

Notes: All the expenditure and income variables are measured in Chinese Yuan (RMB). Telephones include both mobile phones and home phones. Computers include both laptops and desktops. Children includes all children below age 18. Poor health refers to poor self-reported health. Education costs of children refers to the education costs paid for children who study in other cities.

Table 2: Mean of Subsample Variables Used in the Analysis

Variable	Private-sector employees	Junior-level staff BOSOE
Household income	25,187.60	25,924.52
Log of household income	9.94	10.02
Clothing expenditure	1,491.99	1,964.35
Log of clothing expenditure	6.98	7.28
Food expenditure	6,951.80	6,814.89
Log of food expenditure	8.69	8.70
Household equipment and services	1,135.64	1,220.93
Log of household equipment and services	6.29	6.33
No. of cars owned by household	0.01	0.01
No. of motorcycles owned by household	0.25	0.18
No. of telephones owned by household	3.68	3.60
No. of washing machines owned by household	0.98	0.95
No. of televisions own by household	1.27	1.27
No. of fridges owned by household	0.90	0.90
No. of computers owned by household	0.22	0.29
Age of household head	42.02	43.60
No. of children in household	0.68	0.60
Household head has poor health	0.04	0.04
Years of education received by household head	10.40	11.34
Years of education received by spouse	10.09	10.79
Need to pay mortgage	0.04	0.04
Need to pay education costs of children	0.11	0.14
Need to pay rent for accommodation	0.22	0.15
Household head is member of Communist Party	0.21	0.35
Spouse is member of Communist Party	0.12	0.19
Spouse is working	0.77	0.73

Notes: All the expenditure and income variables are measured in Chinese Yuan (RMB). Telephones include both mobile phones and home phones. Computers include both laptops and desktops. Children includes all children below age 18. Poor health refers to poor self-reported health. Education costs of children refers to the education costs paid for children who study in other cities. BOSOE refers to budgetary organizations and state-owned enterprises.

Table 3: ML Estimation Results for the Stochastic Frontier Model

	Coefficient	S.E.
Log of household income	0.962***	0.0431
Assets		
No. of cars owned by household	0.437***	0.156
No. of motorcycles owned by household	–0.102**	0.0410
No. of telephones owned by household	0.0639***	0.0198
No. of washing machines owned by household	0.185***	0.0575
No. of televisions owned by household	0.0770**	0.0353
No. of fridges owned by household	0.238***	0.0511
No. of computers owned by household	0.0566	0.0416
Demographic controls		
Age of household head	–0.0199	0.0202
Square of age of household head	0.000188	0.000229
No. of children in household	–0.0307	0.0382
Household head has poor health	–0.0309	0.0873
Years of education received by household head	–0.00367	0.00702
Years of education received by spouse	0.00144	0.00707
Household controls		
Noncorrupt household head is in budgetary organization	0.0108	0.0479
Noncorrupt spouse is in budgetary organization	–0.0106	0.0571
Need to pay mortgage	0.246***	0.0900
Need to pay education costs for children	–0.0979*	0.0516
Need to pay rent for accommodation	–0.0192	0.0467
Household head is member of Communist Party	0.00253	0.0385
Employer of household head is not at loss	0.0354	0.0535
Spouse is member of Communist Party	–0.0157	0.0444
Household head is a director or manager	–0.0706	0.0531
Spouse is working	0.0661	0.0430
Constant	–2.566***	0.617
$\sigma_{u,i}^2$		
Household with corruption opportunities	–0.253*	0.133
Constant	0.0867	0.115
$\sigma_{v,i}^2$		
Household with corruption opportunities	0.218***	0.0819
Constant	–0.287***	0.0594
N	4,236	

Notes: Dependent variable: logarithm of household expenditure on equipment and services. To save space I do not report the results for the provincial dummies. ***, **, * indicate statistically significant at the 1%, 5% and 10% level, respectively.

Table 4: Estimation Results for the Level of Corruption, by Different Reference Groups and Regions

	Main Results	Reference Group (1)	Reference Group (2)	Reference Group (3)	Rich Provinces	Poor Provinces	Corruption-Worried	Less Worried
N^{nc}	3,381	854	2,527	2,051	1,653	1,728	1,400	1,981
N^{c}	855	891	855	855	437	418	343	512
$\bar{\eta}_i^{nc}$	0.864	0.814	0.879	0.865	0.932	0.804	0.863	0.898
	(0.006)	(0.013)	(0.008)	(0.009)	(0.011)	(0.009)	(0.010)	(0.010)
$\bar{\eta}_i^{c}$	0.761	0.773	0.774	0.799	0.846	0.709	0.879	0.741
	(0.010)	(0.011)	(0.011)	(0.012)	(0.018)	(0.013)	(0.022)	(0.013)
Under-reporting	0.103***	0.042**	0.106***	0.066***	0.086***	0.095***	–0.016	0.156***
	(0.015)	(0.017)	(0.015)	(0.016)	(0.024)	(0.019)	(0.023)	(0.021)
95% Conf. Interval	0.074-0.132	0.009-0.075	0.075-0.136	0.039-0.098	0.039-0.134	0.059-0.132	–0.061-0.029	0.116-0.197
Zhong (2018)	0.155	0.101	0.144	0.114	0.145	0.153	0.040	0.224
Replicated Zhong (2018)	0.146	0.077	0.151	0.149	0.150	0.125	NA	NA

Notes: Standard errors are reported in the parentheses. ***, **, * indicate statistically significant at the 1%, 5% and 10% level, respectively.

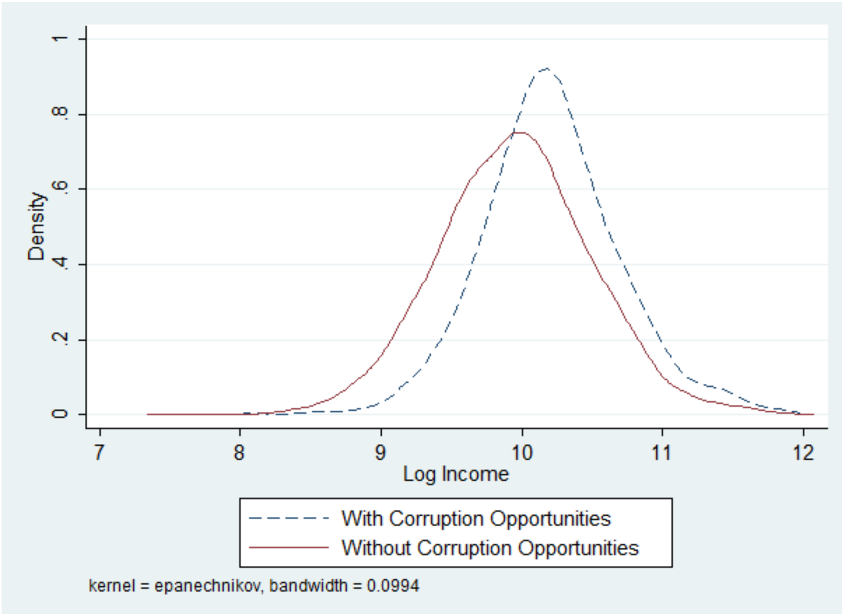


Figure 1: Kernel Density Estimation of the Income Distributions

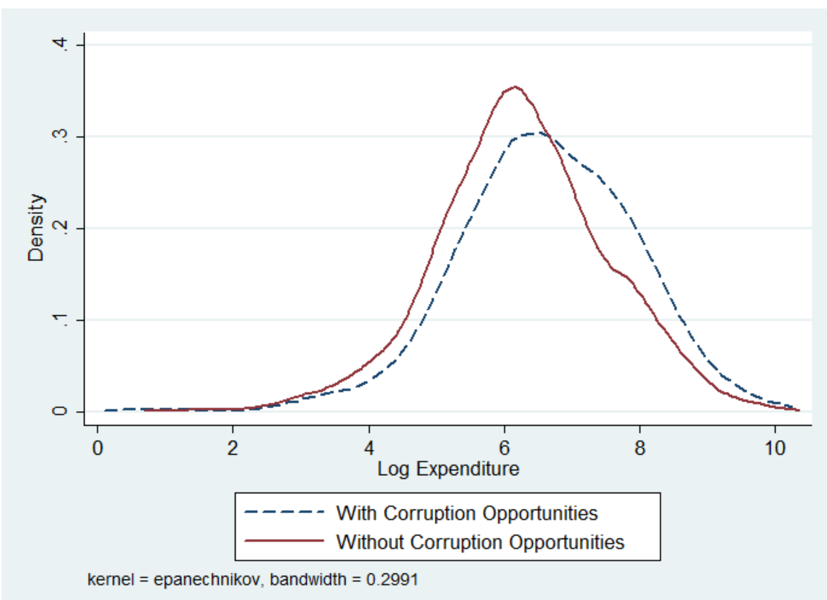


Figure 2: Kernel Density Estimation of the Distributions of the Log of Expenditures on Household Equipment and Services

Appendix A: Failed Replication of Zhong (2018) Using Clothing Expenditure

This appendix records my unsuccessful attempts to replicate certain results of Zhong (2018), who uses the same Chinese household survey as this paper. Zhong borrows from Pissarides and Weber's (1989) method to measure tax evasion. He first uses OLS to estimate an expenditure function for households without opportunities for corruption, then inverts the function to forecast the true income of households with opportunities. The difference between true income and reported income, he assumes, is the value of illegal income. Zhong's most prominent analysis is based on clothing expenditure, although he conducts all his analyses also using expenditure on household equipment and services (the focus of this paper).

As explained in the results section, Zhong uses the method of Pissarides and Weber (1989) when he studies expenditures on equipment and services. In this case he regresses log expenditures on log income, a dummy indicating whether the household is assumed corrupt, and control variables. Zhong, however, cannot use this method when analyzing expenditure on clothing because the clothing expenditure function requires a quadratic term for log income. He thus adapts the method for this special case. In doing so, he must estimate two equations instead of one. The first equation is for noncorrupt households only:

$$\ln C_i = X_i\beta + \gamma_1 \ln Y_i^* + \gamma_2(\ln Y_i^*)^2 + \varepsilon_i, \qquad (AA1)$$

where C_i is clothing expenditure of household i, X is a vector of control variables, β is a vector of parameters, γ_1 and γ_2 are two scalars, ε_i is an error term, and Y_i^* is the household's true income, which in the case of noncorrupt households is their reported income, Y_i. The second equation is for corrupt households only:

$$\ln C_i = X_i\beta + \delta_1 \ln Y_i + \delta_2(\ln Y_i)^2 + \nu_i, \qquad (AA2)$$

where the variables are defined similarly as above and, as Zhong explains, heteroskedasticity must be corrected for at the household level. Zhong shows that the following relationship exists:

$$\delta_1 = \gamma_1 + 2\gamma_2\bar{\mu}, \tag{AA3}$$

where δ_1 is defined in (AA2), γ_1 and γ_2 are defined in (AA3), and $\bar{\mu}$ has a simple explanation that I omit. Zhong then calculates

$$\bar{\theta} = \exp\left(\bar{\mu} + \frac{1}{2}\hat{\sigma}_\eta^2\right), \tag{AA4}$$

which estimates the degree of corruption. The true incomes of the corrupt households are, on average, $(\bar{\theta} - 1)$% higher than their reported incomes. Zhong assumes θ_i (for which $\bar{\theta}$ is the average) is log-normally distributed, which implicitly requires $\theta_i > 1$ (see Zhong's footnote 6). The $\hat{\sigma}_\eta^2$ variable in (AA4), the explanation for which I also omit, is calculated using a simple equation presented in Zhong's appendix.

Table AA1 below presents my estimations of the parameters above, along with Zhong's estimates, for the main sample and the alternative reference groups explained in this paper's results section. I also divide the sample into rich and poor provinces as explained in the results section. Because the method assumes $\theta_i > 1$, whenever my estimation of $\bar{\theta}$ violates this assumption, I refrain from interpretation and instead record "NA" for the underreporting of corrupt households. Similarly, when a calculation requires me to divide by zero, I record "Undefined." Regarding the notation of the parameters in the leftmost column, "Before" indicates the samples used in this paper. "After" indicates samples trimmed in a manner explained below. When I use the same samples used in this paper, I obtain no measurements of underreporting. Regarding the regression for only corrupt households (AA2) I obtain parameter values similar to Zhong's. But regarding the regression for only noncorrupt households (AA1) I obtain starkly different parameter values. After considerable trial and error, I was able somewhat to replicate Zhong's corruption measurement based on clothing expenditure, but only after trimming the noncorrupt group based on the 95th and 1st percentiles of income, dropping 205 households and bringing my noncorrupt-group count closer to Zhong's. This discrepancy could have arisen due to how Zhong and I matched propensity scores. Or Zhong may have trimmed outliers for good reasons. Given that our starting samples differed by 182 households, this seems more likely. (In this paper I choose to include the 205 households. Dropping them leads to only minor differences.) Also, Zhong reports that his propensity scoring led to the dropping of only 85 noncorrupt households from an initial count of 3,334 households. Thus, if the difference in propensity scoring does cause the discrepancy in the parameter values, then apparently Zhong's quadratic method is highly sensitive to preference heterogeneity, which propensity scoring addresses.

Even after the ad-hoc trimming, my measurement of underreporting is only 2.8% compared with Zhong's 18.9%, despite the similarity of our regression coefficients. This underscores the sensitivity of parameter $\bar{\mu}$ of (AA3) and (AA4) to even slight differences in coefficients. In addition, trimming does not produce a robust estimate: I do not obtain reasonable underreporting estimates when using the alternative reference groups, or when I divide by rich and poor provinces. As shown at the bottom of Table 4, I do obtain robust results like Zhong's when I use the one-equation method of Pissarides and Weber (1989) to analyze expenditure on household equipment and services.

A full investigation into why I cannot easily replicate Zhong's clothing results is beyond the scope of this appendix. In general, two-equation methods, especially those in which the sample is divided, are less robust in comparison with one-equation methods. Ideally, I could examine both Zhong's dataset and his computer code. Unfortunately, Zhong says he wrote his paper in 2008 and no longer possesses analysis files or his dataset. He also does not remember precisely how he created it from the raw data. In fact, the only thing he retains is a one-page text file with a few hints. I did my best to recreate his dataset using this text file and the paper itself. I believe I was mostly successful due to the similarity of our summary statistics, as shown in Tables AA2-AA3. And, as mentioned, I am mostly successful in replicating Zhong's results based on expenditures on equipment and services.

Table AA1: Estimation Results for the Level of Corruption, by Different Reference Groups and Regions, Based on Clothing Expenditure

		Main Results	Reference Group (1)	Reference Group (2)	Reference Group (3)	Rich Provinces	Poor Provinces
γ_1	Before	3.918***	5.010***	3.279***	4.079***	1.39*	6.211***
		(0.547)	(1.214)	(0.637)	(0.904)	(0.822)	(0.813)
γ_1	After	5.972***	5.786***	5.488***	7.626***	4.383***	8.374***
		(0.979)	(2.009)	(1.243)	(1.499)	(1.444)	(1.513)
γ_1	Zhong	6.034***	6.011***	6.051***	6.044***	5.969***	6.065***
	(2018)	(0.698)	(1.125)	(1.338)	(1.775)	(0.984)	(1.032)
γ_2	Before	–0.157***	–0.215***	–0.125***	–0.163***	–0.037	–0.269***
		(0.028)	(0.062)	(0.032)	(0.046)	(0.041)	(0.042)
γ_2	After	–0.260***	–0.252**	–0.235***	–0.341***	–0.185**	–0.379***
		(0.050)	(0.104)	(0.063)	(0.077)	(0.072)	(0.078)
γ_2	Zhong	–0.261***	–0.261***	–0.264***	–0.260***	–0.277***	–0.260***
	(2018)	(0.037)	(0.063)	(0.068)	(0.080)	(0.049)	(0.059)
δ_1	Before	5.960***	5.960***	5.960***	5.960***	9.682***	7.934***
		(1.325)	(1.325)	(1.325)	(1.325)	(2.542)	(1.932)
δ_1	After	5.960***	5.960***	5.960***	5.960***	9.682***	7.934***
		(1.325)	(1.325)	(1.325)	(1.325)	(2.542)	(1.932)
δ_1	Zhong	5.948***	5.948***	5.948***	5.948***	5.866***	5.983***
	(2018)	(1.295)	(1.295)	(1.295)	(1.295)	(1.742)	(1.826)
$\bar{\mu}$	Before	–6.503	–2.209	–10.724	–5.770	Undefined	–3.203
$\bar{\mu}$	After	0.023	–0.345	–1.00	2.443	–14.322	0.580
$\bar{\mu}$	Zhong (2018)	0.165	0.121	0.195	0.171	0.186	0.158
$\hat{\sigma}_\eta^2$	Before	0.049	0.039	0.031	0.018	0.071	0.035
$\hat{\sigma}_\eta^2$	After	0.010	0.026	0.023	0.030	0.008	0.004
$\hat{\sigma}_\eta^2$	Zhong (2018)	0.016	0.017	0.014	0.015	0.017	0.016
$\bar{\theta}$	Before	0.002	0.112	0.000	0.003	Undefined	0.041
$\bar{\theta}$	After	1.028	0.979	0.372	11.681	0.000	1.790
$\bar{\theta}$	Zhong (2018)	1.189	1.138	1.224	1.213	1.215	1.180
Underreporting Before		NA	NA	NA	NA	Undefined	NA
Underreporting After		2.8%	NA	NA	1,170%	NA	79.0%
Underreporting Zhong (2018)		18.9%	13.8%	22.4%	21.3%	21.5%	18.0%

Notes: Standard errors are reported in the parentheses. ***, **, * indicate statistically significant at the 1%, 5% and 10% level, respectively.

Table AA2: Mean of Variables Used in the Analysis (Zhong 2018 Values in Parentheses)

Variable	Full Sample	Household with corruption opportunities	Household without corruption opportunities
Household income	25,824.15	30,678.77	24,471.01
	(25,345.07)	(30,575.13)	(24,059.81)
Log of household income	10.01	10.21	9.95
	(9.91)	(10.11)	(9.86)
Clothing expenditure	1,996.86	2,591.22	1,831.19
	(2,002.25)	(2,607.87)	(1,906.47)
Log of clothing expenditure	7.27	7.60	7.18
	(7.24)	(7.59)	(7.15)
Food expenditure	6,807.73	7,384.59	6,646.94
	(6,809.39)	(7,377.87)	(6,663.87)
Log of food expenditure	8.69	8.78	8.67
	(8.67)	(8.76)	(8.65)
Household equipment and services	1,226.66	1,578.66	1,128.55
	(1,178.42)	(1,527.37)	(1,089.09)
Log of household equipment and services	6.34	6.61	6.26
	(6.17)	(6.49)	(6.10)
No. of cars owned by household	0.01	0.01	0.01
	(0.01)	(0.02)	(0.01)
No. of motorcycles owned by household	0.20	0.21	0.19
	(0.21)	(0.22)	(0.20)
No. of telephones owned by household	3.63	3.93	3.55
	(3.26)	(3.50)	(3.19)
No. of washing machines owned by household	0.96	0.99	0.95
	(0.96)	(0.99)	(0.95)
No. of televisions own by household	1.27	1.32	1.26
	(1.25)	(1.31)	(1.24)
No. of fridges owned by household	0.89	0.91	0.89
	(0.88)	(0.90)	(0.87)
No. of computers owned by household	0.27	0.32	0.26
	(0.27)	(0.32)	(0.26)
Age of household head	43.53	44.08	43.38
	(43.08)	(43.44)	(42.99)
No. of children in household	0.61	0.58	0.62
	(0.63)	(0.59)	(0.62)
Household head has poor health	0.04	0.04	0.04
	(0.04)	(0.04)	(0.04)
Years of education received by household head	11.39	12.55	11.06
	(11.35)	(12.60)	(11.11)
Years of education received by spouse	10.76	11.73	10.49
	(10.78)	(11.82)	(10.49)
Need to pay mortgage	0.04	0.06	0.03
	(0.04)	(0.06)	(0.03)
Need to pay education costs of children	0.14	0.16	0.13
	(0.14)	(0.16)	(0.14)

Table AA2 (Cont.)

Variable	Full Sample	Household with corruption opportunities	Household without corruption opportunities
Need to pay rent for accommodation	0.16	0.13	0.17
	(0.16)	(0.13)	(0.17)
Household head is member of Communist Party	0.38	0.62	0.32
	(0.38)	(0.62)	(0.31)
Spouse is member of Communist Party	0.22	0.41	0.17
	(0.23)	(0.42)	(0.18)
Spouse is working	0.72	0.79	0.71
	(0.84)	(0.88)	(0.81)

Notes: All the expenditure and income variables are measured in Chinese Yuan (RMB). Telephones include both mobile phones and home phones. Computers include both laptops and desktops. Children includes all children below age 18. Poor health refers to poor self-reported health. Education costs of children refers to the education costs paid for children who study in other cities.

Table AA2: Mean of Subsample Variables Used in the Analysis (Zhong 2018 Values in Parentheses)

Variable	Private-sector employees	Junior-level staff BOSOE
Household income	25,187.60	25,924.52
	(25,779.08)	(23,045.81)
Log of household income	9.94	10.02
	(9.98)	(9.81)
Clothing expenditure	1,491.99	1,964.35
	(2,097.93)	(1,805.36)
Log of clothing expenditure	6.98	7.28
	(7.29)	(7.10)
Food expenditure	6,951.80	6,814.89
	(6,822.31)	(6,563.77)
Log of food expenditure	8.69	8.70
	(8.69)	(8.61)
Household equipment and services	1,135.64	1,220.93
	(1,202.39)	(1,009.02)
Log of household equipment and services	6.29	6.33
	(6.24)	(6.08)
No. of cars owned by household	0.01	0.01
	(0.02)	(0.01)
No. of motorcycles owned by household	0.25	0.18
	(0.21)	(0.19)

Table AA2 (Cont.)

Variable	Private-sector employees	Junior-level staff BOSOE
No. of telephones owned by household	3.68	3.60
	(3.23)	(3.11)
No. of washing machines owned by household	0.98	0.95
	(0.97)	(0.92)
No. of televisions own by household	1.27	1.27
	(1.26)	(1.22)
No. of fridges owned by household	0.90	0.90
	(0.87)	(0.87)
No. of computers owned by household	0.22	0.29
	(0.26)	(0.25)
Age of household head	42.02	43.60
	(42.91)	(43.05)
No. of children in household	0.68	0.60
	(0.62)	(0.63)
Household head has poor health	0.04	0.04
	(0.04)	(0.04)
Years of education received by household head	10.40	11.34
	(12.08)	(10.76)
Years of education received by spouse	10.09	10.79
	(10.74)	(10.11)
Need to pay mortgage	0.04	0.04
	(0.04)	(0.02)
Need to pay education costs of children	0.11	0.14
	(0.14)	(0.14)
Need to pay rent for accommodation	0.22	0.15
	(0.18)	(0.16)
Household head is member of Communist Party	0.21	0.35
	(0.28)	(0.34)
Spouse is member of Communist Party	0.12	0.19
	(0.17)	(0.19)
Spouse is working	0.77	0.73
	(0.80)	(0.82)

Notes: All the expenditure and income variables are measured in Chinese Yuan (RMB). Telephones include both mobile phones and home phones. Computers include both laptops and desktops. Children includes all children below age 18. Poor health refers to poor self-reported health. Education costs of children refers to the education costs paid for children who study in other cities. BOSOE refers to budgetary organizations and state-owned enterprises.

Appendix B: Estimation Issues

Pre-Estimation: Existence of the One-Sided Error

The likelihood function of an SF model is highly nonlinear, and estimation can be difficult. Given this potential difficulty, it is common in the literature to test the validity of the SF specification prior to the ML estimation (Kumbhakar et al. 2015). If support for the SF specification is unfounded, then time is better spent on considering other specifications rather than on the numerical details of the maximization. Schmidt and Lin (1984) propose an OLS residual test to check for the validity of the model's SF specification. The idea behind the test is that for a production-type SF model with composed error $v_i - u_i, u_i \geq 0$ and v_i distributed symmetrically around zero, the residuals from the corresponding OLS estimation should skew to the left. This is true regardless of the distributional function chosen for u_i in the model estimation after the pretesting. A test of the null hypothesis of no skewness as opposed to the alternative hypothesis can thus be constructed using the OLS residuals. If the estimated skewness has the expected sign, rejection of the null hypothesis provides support for the existence of the one-sided error. For the skewness test Schmidt and Lin (1984) suggest a sample-moment-based statistic, commonly referred to as the $\sqrt{b_1}$ test:

$$\sqrt{b_1} = \frac{m_3}{m_2\sqrt{m_2}} \qquad (AB1)$$

where m_2 and m_2 are the second and third sample moments of the OLS residuals, respectively. The second sample moment of a random variable x is $\sum(x - \bar{x})^2/n$, and the third sample moment is $\sum(x - \bar{x})^3/n$. A result showing $\widehat{\sqrt{b_1}} < 0$ indicates that the OLS residuals are skewed to the left. Under the null hypothesis of no skewness, the statistic should not be statistically different from zero. The distribution of $\sqrt{b_1}$ is nonstandard, and its critical values are tabulated in several studies, e.g., D'Agostino and Pearson (1973).

Coelli (1995) suggests a variant of this test. He notes that under the null hypothesis of no skewness, the third moment of the OLS residuals is asymptotically distributed as a normal random variable with mean 0 and variance $6m_2^3$/N. Thus, the statistic

$$M3T = m_3 / \sqrt{\frac{6m_2^3}{\mathrm{N}}} \qquad (AB2)$$

has an asymptotic distribution of a standard normal random variable.

Turning to my data (see main text), I first conduct an OLS estimation of the expenditure function and plot the histogram of the residuals compared to a normal density. The resulting chart is reproduced in Figure AB1. There appears to be some evidence of a negative skew, although it is far from clear. To formally examine and test this, I use the skewness statistic. The point estimate of the statistic $\sqrt{b_1}$ (AB2) is obtained from the summary statistic of the OLS residuals. The statistic has a value equal to –0.27. The negative sign indicates that the distribution of the residuals skews to the left, which is consistent with a production specification. To assess the statistical significance of the statistic, I conduct the unaltered test as described by D'Agostino, Belanger, and D'Agostino Jr. (1990). The test returns a *p*-value that is less than 0.00; I confidently reject the null hypothesis of no skewness. Furthermore, the MT3 statistic suggested by Coelli (1995), Equation AB2, equals –7.06. Because it has a normal distribution, the critical value is –1.96, so the result confirms the rejection of the null hypothesis of no skewness in the OLS residuals. I have found support for a left-skewed error distribution, and the skewness is statistically significant. I can proceed to the estimation.

Post-Estimation: Existence of the One-Sided Error

Central to the stochastic frontier model is the one-sided error specification, which represents technical inefficiency. It is therefore important to test the existence of the one-sided error for the model. If evidence for the one-sided error specification is not found, the model reduces to a standard regression model for which a simple OLS estimation would suffice. This amounts to a test for the presence of u_i in the model, and a generalized likelihood ratio (LR) test for the null hypothesis of no one-sided error can be constructed based on the log-likelihood values of the OLS (restricted) and the SF (unrestricted) model. Recall that the OLS-residual-based skewness test introduced in the previous section also tests the validity of the one-sided error specification. Although useful as a screening device, the test does not use the information from the distribution functions as the random error. The LR test introduced here is more precise to the specific model I am estimating, but the disadvantage is that it can only be conducted after the ML estimation of the model has been undertaken. The LR test statistic is

$$-2[L(H_0) - L(H_1)], \qquad (AB3)$$

where $L(H_0)$ and $L(H_1)$ are log-likelihood values of the restricted model (OLS) and the unrestricted model (SF), respectively, and the degree of freedom equals the number of restrictions in the test. For the half-normal model, the LR test amounts to testing the hypothesis that $\sigma_u^2 = 0$. The complication of the test is that the null hypothesis of $\sigma_u^2 = 0$ is on the boundary of the parameter value's permissible space, and therefore the LR test statistic does not have a standard chi-square distribution. Coelli (1995) shows that, in such cases, the test has a mixture of chi-square distributions. The critical values of the mixed distribution for hypothesis testing are tabulated in Table 1 of Kodde and Palm (1986).

I test for the presence of u_i in the model by constructing the LR test statistic (AB3) using the log-likelihood values of the OLS and SF models. The statistic is 47.3, which I compare with a 1% critical value of 9.50. I can reject the null hypothesis of no one-sided error.

The Independence of u_i and v_i

A final estimation-related issue is the independence of u_i and v_i. This assumption is not too restrictive for production models in general because v_i represents shocks outside the control of a firm, and therefore it is unlikely to be related to inefficiency, u_i. One can, however, think of cases in which production risk is captured by the v_i term and risk-taking behavior might be reflected in the inefficiency term. Similarly, one can think of shocks to households that may affect the degree to which they underreport income. For example, households in the public sector may embezzle more funds after local disasters. Approaches exist to handle such nonindependence issues but at the cost of making additional assumptions on the correlation between v_i and u_i. Instead of doing this, I follow most of the production literature and assume independence.

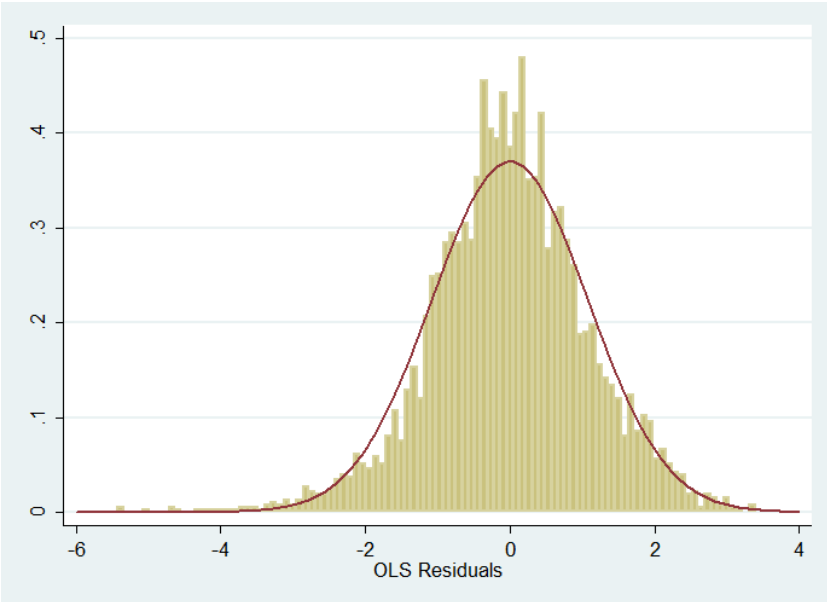


Figure AB1: Histogram of OLS Residuals

4 Chapter 3

Measuring Indonesian Corruption With Household Data: A Stochastic Frontier Approach

Abstract

I apply stochastic frontier (SF) analysis to measure the degree to which public-sector households in Indonesia underreport their income in comparison with private-sector households, assuming the resultant differential is illegal income. I find that the true incomes of public-sector households are, on average, about 50% higher than their reported income, providing a rare measure of corruption's magnitude. I then divide the sample to support the findings of Martinez-Bravo et al. (2017), who exploit the fact that district mayors of the Suharto regime could finish their terms during the democratic transition. Suharto appointed these mayors in 1994, 1995, 1996, or 1997, leading to exogenous variation in corruption exposure. When I restrict my sample to shorter-exposed districts (those with mayors whose appointments were in 1994 or 1995), my corruption measurement falls to 37.1%; when I restrict my sample to longer-exposed districts, my measurement rises to 56.2%. My results provide a correlation absent from Martinez-Bravo et al. (2017): that between mayor exposure and corruption magnitude.

4.1 Introduction

Corruption scholars are always in search of objective measures of bribery and graft. In recent years they have developed a strand of literature in which they compare the results obtained from two datasets: one with corruption and one assumed to be without. The data usually stem from administrative records, special-purpose surveys, or field experiments — all of which can be costly or difficult to obtain. In this paper, I instead use Indonesian household data that is freely available. I apply stochastic frontier (SF) analysis to measure the degree to which public-sector households underreport their income in comparison with private-sector households. I then assume that this differential represents illegal income.

Zhong (2018) is the first to borrow from the tax-evasion literature to impute corruption levels.[1] With data from only a Chinese household survey, he first uses OLS to estimate an expenditure function for households without opportunities for corruption, then inverts the function to forecast the true income of households with opportunities. The difference between true income and reported income, he assumes, is the value of illegal income. Using an expenditure function for household equipment and services, Zhong finds that Chinese households with corruption opportunities underreport their incomes by 14.6%. In Redwine (2018), I measure corruption using the same data but a different method: SF analysis, which is a statistical way to estimate a production frontier and each producer's efficiency based on that frontier. I calculate a "production" frontier in which the output is household expenditure on equipment and services, and the input is household income, controlling for location and household characteristics. I then calculate the degree to which each household "overuses" income in its expenditure. I assume the differential in overuse between household types, those

[1] Specifically, he borrows from Pissarides and Weber's (1989) method to measure tax evasion.

with and without opportunities for corruption, is a measurement of illegal income. In Redwine (2018), I find that Chinese households with opportunities for corruption underreport their incomes by 8.9%, lower than Zhong's estimate but directionally consistent. Furthermore, when I conduct Zhong's robustness checks using the SF method, I detect mostly similar patterns.

The SF method makes two primary assumptions.[2] First, it assumes all households accurately report expenditures on some items, and that bribes in kind or other forms of corruption do not influence these expenditures, which are related solely to the household's true disposable income. Second, some household groups are not corrupt. Alternatively, the method could assume all groups corruptible if the goal is to measure how corrupt one group is compared to another. With these assumptions, I can estimate a reliable expenditure function and uncover illegal income. The SF method, however, has two substantial advantages over Zhong (2018). First, it can test the statistical significance of its measure of corruption, whereas Zhong cannot test the significance of his key parameters, and thus cannot test whether the amount of underreporting is significant.[3] Second, it partially corrects for endogeneity, whereas Zhong does not.

This paper is an extension of Redwine (2018) in that it applies the SF method to another developing country known for corruption: Indonesia. I find that the true incomes of public-sector households are, on average, about 50% higher than their reported income. I also divide my sample in various ways to see the effect on my corruption measurement. Martinez-Bravo et al. (2017) exploit the fact that ex-mayors of the Suharto regime could finish their terms during the democratic transition. The mayors' terms were staggered, so this event led to exogenous variation in corruption exposure. The authors find that districts with longer exposure experience

---

[2] Zhong makes the same assumptions.
[3] Zhong acknowledges this in his paper. In his footnote 17 he says one possible way of estimating standard errors might be to design an appropriate bootstrapping strategy.

worse government outcomes, higher elite persistence, and lower political competition. My results reinforce their findings in a novel way by addressing corruption magnitude: when I restrict my sample to shorter-exposed districts, my corruption measurement falls to 37.1%. When I restrict my sample to longer-exposed districts, my measurement rises to 56.2%. I also divide my sample based on poor and rich districts, finding corruption much heavier in poor areas. Finally, most corruption studies rely on perception-based corruption indices, which use corruption ratings based on expert opinions or surveys of business executives. Several studies, however, have found that perceptions of corruption are not perfect measures of actual corruption (e.g., Olken 2009 and Donchev and Ujhelyi 2014). This paper adds to this literature by exploiting a corruption-related question in the household survey. Using the SF method to measure corruption, I find that households that report trust for the police live in areas with heavier corruption, perhaps because the least informed citizens live in the most corrupt areas — in which officials and the press do little to expose that corruption to the community.

Gorodnichenko and Peter (2007) were the first to measure corruption using only household data. They found that public employees in the Ukraine were underpaid compared with their private-sector counterparts. The wage gap remained even after they controlled for observable and unobservable characteristics, corrected for endogeneity, and accounted for differences in working hours, satisfaction, fringe benefits, job security, bonuses, and secondary employment. At the same time, the levels of expenditures and asset holdings reported by officials and their private-sector counterparts were the same, indicating bribery as long as the labor market was free to reach an equilibrium in which total worker compensation was equalized across sectors. Gorodnichenko and Peter provided evidence that no mobility constraints existed. Because

Indonesian public employees are paid more, on average, than their private counterparts, I cannot apply Gorodnichenko and Peter's method.

To the best of my knowledge, only two other papers measure corruption using household data. Both rely on methods different from the one used in this paper. Saha et al. (2014) analyze the private-public wage gap in India to impute the existence of the unreported income. Because the gap is both positive and negative, depending on income level, the authors use a quantile regression technique. Their hypothesis (for which they find support) is that if they observe positive gaps for both wages and nondurable-good expenditures, but no such gap for durable-good expenditures, then they provide evidence of unreported earnings for official households. Nguyen (2017) examines whether natural disasters in Vietnam have the same effect on the income differential between official and nonofficial households as it does on the expenditure differential. Relying on the exogeneous effects of disasters on income and expenditure, Nguyen runs a series of regressions with panel data and fixed effects. Included in his analysis is a regression for disaster-affected areas only. The dependent variable is the ratio of expenditure to income, logged to represent the expenditure-income gap. Among the independent variables is an indicator of whether the household contains an official. The coefficient of this variable is positive and significant, implying that the official households in the affected areas spend more, even after controlling for reported income. Nguyen assumes this extra expenditure is funded by illegal income.

Also to the best of my knowledge, few micro-level papers have employed SF analysis to uncover the underreporting of income. A noteworthy example is Shonkwiler et al. (2011). They use SF analysis to measure the underreporting of remittances in Armenia. After controlling for this, they find a strong negative impact of remittances on incentives to work.

4.2 Method

4.2.1 Overview

As discussed above, I estimate a production function in which the output is household expenditure on certain items, and the input is household income, with controls for location and household characteristics. I then calculate the degree to which firms "overuse" income. I hypothesize that in comparison with private-sector households, public-sector households seem to require less *reported* income to buy the same amount of goods (i.e., they are more "efficient") — but only because they do not report their illegal income. That is, I assume the differential in overuse between the two household types is a measurement of the illegal income obtained through corruption such as bribery and graft.

The SF approach requires that households accurately report expenditure for at least one item. Disposable income must determine this expenditure, meaning bribes in kind must not include this item. The data sample of this study is from the fourth wave of the Indonesian Family Life Survey (IFLS4). The survey includes household expenditures on a wide variety of items, of which, I argue, true disposable income most heavily influences expenditures on clothing, household supplies, and furniture. Thus "household expenditure" in this paper refers only to the sum of these expenditures. Favor-seekers hardly ever use clothing as bribes because of the difficulty in obtaining accurate information about the recipient's size and tastes. Furthermore, households are unlikely to claim clothing expenditure as a tax-deductible business expense, and employers are unlikely to subsidize it. Likewise, people are unlikely to bribe with household supplies and furniture. Reports issued by the United Nations Office on Drugs and Crime (UNODC) empirically support these assumptions.[4]

SF analysis estimates production inefficiency.[5] There are two possible measures of inefficiency: output-oriented (OO) and input-oriented (IO). The IO measure starts from the fact that if a producer is not efficient, it does not use inputs effectively. That is, there are slacks in the inputs, and producers can reduce input usage without reducing output. Consequently, the IO measure is practical and intuitive when output is exogenous and the objective is to maximize the proportional reduction in input usage without changing output. This potential maximum reduction — measured as a percentage of all inputs — is the IO measure. By contrast, OO technical inefficiency measures the potential increase in output without increasing the input quantities. Alternatively, it can be viewed as a measure of output loss resulting from failure to produce the maximum possible output permitted by the technology. Thus, the OO measure is intuitive when the inputs are exogenously given, and the objective is to produce as much output as possible.

To summarize, a production plan is technically inefficient if a higher level of output is attainable for the given inputs (OO measure), or the observed output level can be produced using fewer inputs (IO measure). In my context I assume household expenditure is a response to household income, so I choose the OO measure of inefficiency, assuming income is exogenous. Thus expenditure is the dependent variable, and income is the independent variable. But I do not wish to measure the magnitude by which households could increase their expenditure given their income (the OO measure). I wish to measure each household's overuse of income given its

[4] See Zhong (2018) for a brief description of these reports.
[5] Kumbhakar et al. (2015) provide a good description of SF analysis.

expenditure choice. This second type of IO measure is subtly different from the one already described, which would require income to be the dependent variable. That is, this second type of IO measure is defined as the first type but estimated differently, leading to a different numerical value.[6] Fortunately, as I explain below, I can easily derive this second type of IO measure from the OO measure.

Estimates of inefficiency are conditional on the given technology (production frontier). An input-output combination may appear inefficient for one technology, but it could be efficient with respect to a different technology. The implication for empirical analysis is that when estimating the technical inefficiencies of different producers, it is important that they are estimated with respect to the appropriate technology. In my application, technology is expenditure preferences: both types of households must have similar preferences for both clothing and household items. In the next section, I partly control for possible differences through propensity scoring.

4.2.2. Functional Form

To model the production function, I use the Cobb-Douglas functional form:

$$y = f(x) = A\prod_{j=1}^{J} x_j^{\beta_j}$$

$$\Rightarrow \quad \ln y = \beta_0 + \sum_j \beta_j \ln x_j \, , \tag{1}$$

where $\beta_0 = A$.

[6] Kumbhakar et al. (2015) note that this issue is like regressing y on x and x on y, in which case the estimated slope coefficients are not reciprocals.

In my application, income is the only input, so the functional forms simplify to

$$y = f(x) = Ax_1^{\beta_1}$$

$$\Rightarrow \quad \ln y = \beta_0 + \beta_1 \ln x\,. \tag{2}$$

OO technical inefficiency enters the production function as the u term in the following expression:

$$y = f(x)e^{-u}. \tag{3}$$

The logarithm of the production function makes $-u$ an additive term to the corresponding neoclassical specification

$$\ln y = \ln f(x) - u. \tag{4}$$

With IO technical inefficiency (η) in the production function, the expression is

$$y = f(xe^{-\eta}), \tag{5}$$

which, in logarithmic form, generally becomes

$$\ln y = \beta_0 + \sum \beta_j \ln x_j - \left(\sum_j \beta_j\right)\eta, \tag{6}$$

which is essentially the same as the OO model with the reparameterization

$$u = \eta \sum\nolimits_j \beta_j\,. \tag{7}$$

Thus, once I estimate u, I can easily obtain η from the relationship expressed in (7). In my application this means dividing u by the coefficient of log income.[7] In general terms, producers on average use $(\eta \times 100)\%$ more input than necessary due to technical inefficiency. I hypothesize that households with corruption opportunities overuse income to a lesser degree only because they do not report their illegal income. So my method for measuring illegal income consists of estimating u_i for each household i (using the method described below), deriving an observation-specific η_i from u_i, then calculating the average η_i for both types of households. The difference between these averages is my measurement of illegal income as a percentage of reported income.

As already mentioned, I assume that income is exogenous in the expenditure function, even though it is endogenous. Nevertheless, the SF method addresses this endogeneity by construction: the corruption measurement is the *difference* in two measurements of overuse (the two η_i averages), each constructed using the same (likely biased) coefficient of log income. If the two overuse measures are biased similarly due to this biased coefficient, the corruption measure should be mostly unbiased due to income endogeneity (though, as I discuss later, I must correct for possible bias in u_i itself due to heteroskedasticity). In the results section I report the overuse of income for each household type, but due to their biases, I analyze only their differences.

[7] The intuition is as follows: think of a single y and single x. Because the relationship between $\ln y$ and $\ln x$ is linear, the vertical distance from a point below the line (which measures OO inefficiency, u) is the product of the horizontal distance (IO inefficiency, η) times the slope (β), that is, $u = \eta\beta$.

4.2.3 Estimation of Mean Inefficiency

To estimate OO technical efficiency, I can specify a stochastic production frontier model with OO technical efficiency as

$$\ln y_i = \ln y_i^* - u_i, \qquad u_i \geq 0, \tag{8}$$

$$\ln y_i^* = f(\boldsymbol{x}_i; \boldsymbol{\beta}) + v_i, \tag{9}$$

where the subscript *i* denotes observations (households), y_i is a scalar of observed output (expenditure), $\boldsymbol{x}_i$ is a J×1 vector of the one input variable (income) and the controls, $\boldsymbol{\beta}$ is a vector of the corresponding coefficient vector, v_i is a zero-mean random error, and $u_i \geq 0$ is production inefficiency. Equation (9) defines the stochastic production frontier function. Given $\boldsymbol{x}$, the frontier gives the maximum possible level of output, and it is stochastic because of v_i. Given that $u_i \geq 0$, observed output (y_i) is bounded below the frontier output level (y_i^*). I also can write the model in the form:

$$\ln y_i = f(\boldsymbol{x}_i; \boldsymbol{\beta}) + \epsilon_i, \tag{10}$$

$$\epsilon_i = v_i - u_i, \tag{11}$$

where ϵ_i is the error term often called the composed error term.

I construct a statistic to estimate u_i as follows. First, recall that I assume the Cobb-Douglas functional form for the production frontier $f(x)$. The estimation of the model then involves (i) estimating the parameters of the frontier function $f(x)$, and (ii) estimating inefficiency. To estimate $f(x)$ I first impose distributional assumptions on the error components, then derive the log-likelihood function of the model and use numerical maximization procedures to obtain the ML estimates of the model parameters. A zero-mean normal distribution for v_i is widely accepted in this context (Kumbhaker et al. 2015). The choice of distributional assumption for the random variable u_i is more the issue at stake. The distribution must be in the nonnegative domain, and its joint distribution with v_i would ideally have a closed form. The literature has identified several such distributions. In the first estimation of the SF model with distributional assumptions on v_i and u_i, Aigner et al. (1977) assumed a half-normal distribution for u_i. The half-normal distribution has a single parameter and is thus relatively easy to estimate. Subsequent developments in the literature have suggested more flexible (but harder to estimate) distribution functions such as the truncated-normal distribution with or without scaling properties (Stevenson 1980; Wang and Schmidt 2002). For my application, however, a half-normal distribution is appropriate. Recall that I estimate u_i and then divide by a scalar to obtain η_i, which measures how households overuse income. Households seem to overuse income because other households do not report their full income and thereby seem "efficient" to varying degrees. In this way the most dishonest households define the frontier. Studies have shown that survey respondents tend to underreport income (Ravallion 2003; Freund and Spatafora 2008). I therefore expect most households to be efficient, which would result in a clustering of η_i near a calibrated zero, with a tail to the right representing the households that report honestly to varying degrees.[8]

Based on (8) and (9), a production SF with a normal distribution on v_i and a half-normal distribution on u_i is represented as the following:

$$\ln y_i = \ln y_i^* - u_i, \tag{12}$$

$$\ln y_i^* = \boldsymbol{x_i \beta} + v_i, \tag{13}$$

$$u_i \sim i.i.d.\, N^+(0, \sigma_u^2), \tag{14}$$

[8] The shadow economy could exacerbate underreporting. According to Schneider et al. (2011), Indonesia's shadow economy in 2005 was 19.1% of GDP.

$$v_i \sim i.i.d. N(0, \sigma_v^2), \tag{15}$$

where v_i and u_i are distributed independently of each other. The $\boldsymbol{\beta}$, σ_u^2, and σ_v^2 are the parameters to be estimated. Equation (14) assumes that the inefficiency effect follows a half-normal distribution. One way to derive the half-normal distribution is to treat it as the nonnegative truncation of a zero-mean normal distribution. I shall denote the distribution derived in this way as $N^+(0, \sigma_u^2)$, where σ_u^2 is the variance of the normal distribution before truncation. Suppose that a random variable Z has a normal distribution $z \sim N(\mu, \sigma_z^2)$ with the probability density function denoted by $g(z)$. If it is truncated from above at the point α so that $z \geq \alpha$, then the density function of z, $f(z)$, is

$$f(z) = \frac{g(z)}{1 - \Phi\left(\frac{\alpha - \mu}{\sigma_z}\right)} = \frac{\frac{1}{\sigma_z}\phi\left(\frac{z - \mu}{\sigma_z}\right)}{1 - \Phi\left(\frac{\alpha - \mu}{\sigma_z}\right)}, \qquad z \geq \alpha, \tag{16}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the probability density and probability distribution functions, respectively, for the standard normal variable.[9] The density function of u_i in (14) can then be obtained by setting $\mu = 0$ and $\alpha = 0$ in the above equation to give the following:

$$f(u_i) = \frac{\frac{1}{\sigma}\phi\left(\frac{u_i}{\sigma}\right)}{1 - \Phi(0)} = \frac{2}{\sigma}\phi\left(\frac{u_i}{\sigma}\right) = 2(2\pi\sigma^2)^{-\frac{1}{2}}\exp\left(-\frac{u_i^2}{2\sigma^2}\right), \qquad u_i \geq 0. \tag{17}$$

The log-likelihood function based on (12)-(15) for each observation i is

$$L_i = -\ln\left(\frac{1}{2}\right) - \frac{1}{2}\ln(\sigma_v^2 + \sigma_u^2) + \ln\phi\left(\frac{\epsilon_i}{\sqrt{\sigma_v^2 + \sigma_u^2}}\right) + \ln\Phi\left(\frac{\mu_{*i}}{\sigma_*}\right), \tag{18}$$

[9] See Johnson et al. (1995).

where

$$\mu_{*i} = \frac{-\sigma_u^2 \epsilon_i}{\sigma_v^2 + \sigma_u^2}, \tag{19}$$

$$\sigma_*^2 = \frac{\sigma_v^2 \sigma_u^2}{\sigma_v^2 + \sigma_u^2}. \tag{20}$$

For detailed derivations, see Appendix A of Kumbhakar et al. (2015). The log-likelihood function is then the observational sum of (18), which can then be numerically maximized to obtain the estimates of the model parameters. There is, however, a computational problem. The variance parameters, σ_u^2 and σ_v^2, must be positive, but an unconstrained numerical maximization would not guarantee positive estimates. To ensure that the variance parameter estimates are positive, researchers in the early literature often used the following parameterization scheme for the unconstrained numerical maximization:

$$\sigma_u^2 = \exp(w_u), \tag{21}$$

$$\sigma_v^2 = \exp(w_v), \tag{22}$$

where w_u and w_v are unrestricted constant parameters. (Kumbhakar et al. 2015). As I explain momentarily, I choose to go step further and correct of heteroskedasticity.

4.2.4 Correcting for Heteroskedasticity

The original half-normal model of Aigner et al. (1977) assumes that the v_i and the pretruncated u_i are homoskedastic, that is, both σ_v^2 and σ_u^2 are constants. Caudill and Ford (1993), Caudill, Ford, and Gropper (1995), and Hadri (1999) consider models in which these random variables are heteroskedastic. Unlike a classical linear model in which heteroskedasticity affects only the efficiency of the estimators and not their consistency, ignoring heteroskedasticity in the SF framework leads to inconsistent estimates (Wang and Schmidt 2002). Kumbhakar and Lovell (2000, Section 3.4) provide a detailed discussion on the consequences of ignoring the heteroskedasticity, assuming v_i and u_i are heteroskedastic. Ignoring the heteroskedasticity of v_i still gives consistent estimates of the frontier function parameters ($\boldsymbol{\beta}$) except for the intercept, which is downward-biased. Estimates of the technical efficiency are biased. Ignoring the heteroskedasticity of u_i causes biased estimates of the frontier function's parameters as well as the estimates of technical efficiency. Caudill and Ford (1993), Caudill, Ford, and Gropper (1995), and Hadri (1999) propose the heteroskedasticity can be parameterized by a vector of observable variables and associated parameters. For instance, $\sigma_{u,i}^2 = \exp\left(\boldsymbol{z}_{u,i}; \boldsymbol{w}_u\right)$, where $\boldsymbol{z}_{u,i}$ is an $m \times 1$ vector of variables including a constant of 1, and $\boldsymbol{w}_u$ is the $m \times 1$ corresponding parameter vector. The exponential function is used to ensure a positive estimate of the variance parameter. Therefore, the parameterizations are

$$\sigma_{u,i}^2 = \exp\left(\boldsymbol{z}'_{u,i}\boldsymbol{w}_u\right), \tag{23}$$

$$\sigma_{v,i}^2 = \exp\left(\boldsymbol{z}'_{v,i}\boldsymbol{w}_v\right). \tag{24}$$

The vectors $\boldsymbol{z}_{u,i}$ and $\boldsymbol{z}_{v,i}$ may or may not be the same vector, and they may contain all or part of the $\boldsymbol{x}_i$ vector. In my application I am interested in the difference between the mean u_i (divided by the coefficient of log income) for households with opportunities for corruption and the mean u_i (divided by the same scalar) for all other households. My main concern, therefore, is how the bias in u_i due to heteroskedasticity differs between household types, so I parameterize both $\sigma_{u,i}^2$ and $\sigma_{v,i}^2$ with an indicator variable for households with opportunities for corruption.

The log-likelihood function of the heteroskedastic model is the same as in (18), except that I now use (23) and (24) instead of (21) and (22) in place of σ_u^2 and σ_v^2, respectively, in the log-likelihood function. All the parameters of the model are estimated at the same time via the ML method.

4.2.5 Estimation of Observation-Specific u_i

After I estimate the model parameters, I can estimate the observation-specific u_i. Although the definition of this index is intuitive, estimating the index for each observation is less straightforward. To see this, note that $u_i \sim N^+(0, \sigma_u^2)$. The ML estimation of the model yields the estimated value of σ_u^2, which provides information about the shape of the half-normal distribution of u_i. This information is all I need to find the average technical inefficiency of the sample. However, I am interested in the u_i of each observation, so this information on σ_u^2 is not enough because it does not contain any household-specific information. The solution, first proposed by Jondrow et al. (1982), is to estimate u_i from the expected value of u_i conditional on the composed error of the model, $\epsilon_i \equiv v_i - u_i$. This conditional mean of u_i given ϵ_i gives a point estimate of u_i. The composed error contains household-specific information, and so the conditional expectation yields the observation-specific value of the inefficiency.[10] Jondrow et al. (1982) show that the density function of $(u_i|\epsilon_i)$ is $N^+(u_{*i}, \sigma_*^2)$, based on which, the equation of $E(u_i|\epsilon_i)$ is (see Appendix B of Kumbhakar et al. 2015):

$$E(u_i|\epsilon_i) = \frac{\sigma_* \phi\left(\frac{\mu_{*i}}{\sigma_*}\right)}{\Phi\left(\frac{\mu_{*i}}{\sigma_*}\right)} + \mu_{*i}, \tag{25}$$

[10] Kumbhakar et al. (2015) likens this to extracting signal from noise.

where μ_{*i} and σ_* are defined in (19) and (20). I substitute into the equation ML estimates of the parameters to obtain the empirical estimate of inefficiency, which is guaranteed to be nonnegative. Estimates of observation-specific inefficiency assume the model parameters are known and given, while in fact I estimate them with uncertainty.

4.3 Data and Variables

The IFLS is an ongoing longitudinal survey in Indonesia. The sample is representative of about 83% of the population and contains more than 30,000 people living in 13 of the country's 27 provinces. This study primarily uses data from the fourth wave (IFLS4), which was conducted in 2007 and 2008 by RAND, the Center for Population and Policy Studies of the University of Gadjah Mada, and Survey METRE. One question in the IFLS4 asked respondents, "Which category best describes the work you do?" The answer options included self-employed (in various categories), government worker, private worker, casual worker in agriculture, casual worker not in agriculture, and unpaid family worker. I selected my sample according to three criteria: (i) the household contains at least two adults; (ii) one adult is the spouse; and (iii) the head of household is either a private worker or a government worker (excluding the self-employed, casual and unpaid workers, and retired people). The purpose of this sample selection process is to avoid differences owing to the preferences of various other household groups, such as single parents, the self-employed, and pensioners. In my analysis, I start with the assumption that household members who do not work for the government cannot engage in corruption, but that all government workers can do so. These assumptions may lead to an underestimation of the extent of corruption if private workers engage in corrupt activities. In this case, the estimate of corruption serves as a lower bound. Alternatively, I can consider my corruption estimate as a measure relative to the reference group. I define households with one or more government workers as public-sector households with opportunities for corruption, and all other households as private-sector households without them. I classify 601 observations as public-sector households, and 1,626 observations as private-sector households.

I calculate illegal income by estimating how public-sector households differ in their underreporting of income when compared with private-sector households. If the two types of households differ in their underreporting for another reason — tax evasion — then my results are confounded. To address this concern, I exclude all self-employed respondents from my sample. For wage earners, it is relatively difficult to underreport employment income because of employer reporting. Moreover, unlike most developed countries, the personal income tax is not a major source of government revenue in Indonesia.[11] I therefore believe that tax-evasion motives affect my results only negligibly, although I cannot test this using the IFLS data.

4.3.1 Variables

The dependent variable is the logarithm of household expenditure on clothing, household supplies, and furniture. The explanatory variable is the logarithm of household income, which is the sum of a household's cash income and other incomes in kind from all sources, minus taxes and fees. All expenditure and income variables are measured in Indonesian rupiah (Rp). Dummy control variables linked to housing characteristics include an inside water source, an outside kitchen, wood floors, and a moderately sized yard. Control variables linked to assets include four dummies indicating whether the household owns (separately) a toilet, fridge, stove, and television. I also control for the number of chickens and ducks owned by the household. Demographic control variables include the age of the head of household, the square of the age of

---

[11] Receipts from taxes on income, profits, and capital gains of individuals accounted for just 13.6% of Indonesia's total revenue in 2008 (OECD 2008).

the head of household, the number of children in the household, an indicator whether the head of household is in poor health, and years of education attained by the head of household and the spouse. Control variables concerning household characteristics include indicators for whether the household owns its home, the household pays education costs, the household pays rent for accommodation, the household is in an urban area, the head of household is Muslim, and the spouse is working. Also included is a set of dummy variables for the district in which the household resides. Table 1 presents the summary statistics for all variables for the whole sample, public-sector households, and private-sector households. The statistics indicate that the characteristics of the two types of households are different. In terms of income, consumptions, and household assets, public-sector households are considerably better off. My method relies on the relationship between income and expenditures, so the difference between the income distributions is a concern. Correct results clearly depend on getting Engel effects right, the more so as the income distributions of the two household types differ. I draw kernel density plots of these two income distributions in Figure 1, which shows that the income distribution for public-sector households dominates that for private-sector households.[12] However, the ranges of the two distributions look similar. In this case, if I can ensure that the two types of households are similar in preferences, then the estimated income-expenditure relationship can be used to fully reflect the true level of corruption. Figure 2 presents kernel density plots of distributions of expenditures for the two groups. The expenditure distribution for public-sector households dominates that for the private-sector households.

[12] Kolmogorov-Smirnov tests support my comparisons at the 1% level regarding both income distributions and expenditure distributions of the two household types.

4.4 Empirical Results

4.4.1 Accounting for Preference Heterogeneity

My analysis has not yet accounted for preference heterogeneity. If public-sector households, compared to private-sector households, have different preferences for clothing and household items, then my measurement of illegal income will capture both preference heterogeneity and corruption. For example, officials might enjoy more generous fringe benefits and added job security, and these things could lead to increased spending on luxury clothing. More generally, people with different levels of income, wealth, education, household characteristics, location of residence, and other socioeconomic characteristics may have different consumption preferences. If some public-sector households, compared to private-sector households, are very different in those factors, then my corruption interpretation could be incorrect. To mitigate this possibility, I use propensity scoring to restrict those factors in a common support. I calculate the propensity score for each household and then restrict the sample to areas of common support. I estimate a probit model in which the outcome variable is an indicator of whether the household is in the public sector. Independent variables include log of income and some other variables that may affect household consumption.[13] By excluding those observations that are not in the region of common support, I drop 60 households.

4.4.2 Main Results

Before I estimate illegal income, I follow the SF literature by testing for the one-sided error term,

[13] See Zhong (2018) for a more thorough explanation of propensity-score matching in a similar context. Independent variables in the probit model include log income and indicators of an inside water source, an outside kitchen, wood floors, a moderately sized yard, the household owns its home, the household pays education costs, the household pays rent for accommodation, the household is in an urban area, the head of household is Muslim, the spouse is working, and the ownership of a toilet, fridge, stove, and television. Also included are the number of chickens and ducks owned by the household, as well as the age of the head of household, the number of children in the household, and years of education attained by the head of household and the spouse.

as explained in Appendix A. The pre-estimation tests (Equations A1-A2) support the existence of a one-sided error term, as does a post-estimation test (A3).

Using the method outlined in the previous section I proceed to estimate the model of Equations (12)-(15) with parametrical distributional assumptions on v_i and u_i, assuming normal and half-normal distributions, respectively, and correcting for heteroskedasticity. See Table 2 for the ML estimation. The coefficients of the indicator variable for a public-sector household in both $\sigma_{u,i}^2$ and $\sigma_{v,i}^2$ are statistically significant, implying that the inclusion of this variable in the model is supported by the data (See Equations [23]-[24] and the related discussion).

Next, I use Equation 25 to generate the estimated values of $E(u_i|\epsilon_i)$ evaluated at $\hat{\epsilon}_i$.[14] I divide these values by the coefficient of log income to obtain the observation-specific η_i, averages of which I calculate for both types of households. Table 3 reports these averages, denoted $\bar{\eta}_i^{private}$ and $\bar{\eta}_i^{public}$, for the noncorrupt and corrupt households, respectively, along with all figures discussed below.[15] The difference between these averages is the illegal income of households with opportunities for corruption, expressed as a percentage of reported income. This estimate is 49.7% with a 95% confidence interval of 42.5% to 57.0%. Recall that this illegal income is specifically from bribery and graft under the assumptions that public-sector and private-sector households hide income identically for other reasons such as tax evasion and black-market participation. Also recall that I can interpret this result as a corruption measure for public-sector households in relation to corruption of private-sector households. As indicated in Table 3, all

[14] To estimate the model, I first set up the likelihood function by using the *sfmodel* Stata command written by the authors of Kumbhakar et al. (2015). I then estimate the model by using Stata's *ml max* command. Finally, I generate the observation-specific $E(u_i|\epsilon_i)$ using the *sf_search* command, also provided by Kumbhakar et al. (2015).

[15] As discussed in the method section, one should interpret the η_i averages with care because they are potentially very biased.

estimates of underreporting mentioned in this paper are significant at the 1% level according to two-sided *t*-tests.

4.4.3 Estimations for Different Groups and Regions

How democratic transitions unfold, Martinez-Bravo et al. (2017) show, is a key determinant of the extent of elite capture (interest groups' influence over politics). The authors exploit quasi-random variation that originated during the Indonesian transition: Suharto-regime mayors could finish their five-year terms before being replaced by new leaders. Because mayors' political cycles were not synchronized, this event generated exogenous variation in how long old-regime mayors remained in their position during the democratic transition. The authors find that slower transitions toward democracy allow the old-regime elites to capture democracy. Districts with longer exposure experience worse government outcomes, higher elite persistence, and lower political competition in the medium run. Most relevant to this study is the finding that private firms in longer-exposed districts are more likely to report that they face regular extortion from the military and police groups — a finding that sheds light on corruption's extensive margin. To test whether my corruption measurement corroborates this story for the intensive margin, I obtained certain data from the authors (Martinez-Bravo and Stegmann 2018) that allow me to divide my sample based on whether the district mayors were appointed either in 1994 or 1995, or in 1996 or 1997. The districts with the 1996 and 1997 appointments experienced longer exposure to Suharto-appointed mayors. When I drop these districts from my sample (leaving 52.3% of the sample), average underreporting is 37.1%. In Table 3 I label this group "Shorter Exposure." When I consider only the opposite group, labelled "Longer Exposure," average underreporting is 56.2%. The 95% confidence interval of the latter measure is narrow — 55.4% to 56.9% — and its lower bound is above the higher bound of the corresponding 95% confidence interval for shorter-exposed districts. Thus I find a strong treatment effect that is statistically significant, providing further evidence that Suharto-appointed mayors helped maintain corruption.

The IFLS4 survey asks respondents to imagine that a police officer has found their lost wallet or purse containing their ID and 200,000 rupiah (about $14 at the time). The survey then asks whether it is likely or unlikely that the officer will return the wallet with the money in it. When I drop from my sample people who distrust the police (leaving 63.9% of the sample), average underreporting is 76.7%. In Table 3 this group is labelled "Trust Police." When I consider only the opposite group, labelled "Distrust Police," average underreporting is 33.8%. So the households who trust the police experience greater corruption. One can only speculate about the reasons, as researchers have done with similar findings (see citations in the introduction). A greater concern for corruption perhaps leads to less corruption, or maybe more corrupt areas contain more corrupt respondents, who may be less inclined to indicate distrust for the police.[16] Regardless, my finding is further evidence that researchers should be wary of relying too heavily on attitudinal survey questions when assessing the extent of corruption.[17]

Finally, I divide the sample into two groups: rich districts and poor districts.[18] I then estimate the level of corruption for each group. Underreporting is 71.4% for poor districts and 31.0% for rich districts. Because of the bidirectional relationship between economic development and corruption, these results were not a foregone conclusion. Corruption may harm economic

[16] This latter conjecture seems possible because the proportion of households that trust the police varies by household type: 74.5% of public-sector households, but only 60.4% of private-sector households.
[17] Similarly, using this paper's method to measure corruption, Redwine (2018) finds that Chinese households that report the most concern for corruption experience the least corruption.
[18] I rank the 37 districts by average household income, calculated from the IFLS4 data. I then define the top half of districts as rich and the other half as poor.

development, but a high level of economic development may provide more opportunities for corruption. My results suggest that the former effect dominates in Indonesia.[19]

4.5 Conclusion

This paper provides a rarity in the literature: an objective estimate of the extent of corruption, in this case for Indonesia in 2008. According to the SF method, the true incomes of public-sector households are, on average, about 50% higher than their reported income. Recall that this estimate depends on the relatively strong assumption that private workers are not hiding income derived from bribery or graft. This paper's corruption measures, therefore, are best thought of either as lower-bound estimations or as relative measures of corruption (my preference).

Using the same SF method with household data, Redwine (2018) finds that Chinese officials underreport their income by only 10.3%. The reference group here includes not only private workers, but also officials with fewer opportunities for corruption. When the reference group is only private workers, the corruption estimate lowers to 4.2%. Why does the SF method produce such widely different corruption estimates for China and Indonesia? Likely explanations include differences in the household surveys. The Chinese survey, for example, includes only urban households. Furthermore, the conductors of the surveys might have addressed, with different success, a particularly relevant form of self-selection: the most corrupt public officials may have been less inclined to participate. Another reason for the disparity in corruption measurements could be the differences in the societies: two rapidly growing economies that have evolved, on the one hand, from rigid Chinese planning and, on the other, from Suharto-appointed monopolies. National statistics imply that the public sector permeates China to a greater

[19] By contrast, Zhong (2018) finds that the latter effect dominates in China, though Redwine (2018) fails to replicate this result.

degree.[20] Perhaps this saturation has led to a more corrupt Chinese private sector, lowering the relative estimate of corruption for public workers. Other possible forces behind the corruption-magnitude discrepancy between China and Indonesia include differing public-private differentials in relation to propensities to save and black-market participation, both of which could be confounding my results. And, of course, Indonesia simply may be more corrupt: the country does typically fare worse than China in corruption-perception studies.[21]

Regarding future research, the IFLS is a longitudinal study, so possibly I could employ the SF method to study corruption estimates across time, using the years 1993/1994, 1997, 2000, 2008/2009 (the data of this study), and 2014/2015. A simple approach is to compare cross-sectional measurements.[22] In a quick first attempt, I organized the 2014/2015 data as I did for this paper, then ran the resultant data through the same computer code.[23] Surprisingly, in comparison with the 2008/2009 analysis, the number of public-sector households drops from 541 to 371, and the number of private-sector households drops from 1,626 to 1,187. Nevertheless, I find that the true incomes of public-sector households are, on average, 55.9% higher than their reported incomes, a figure that is significant at the 1% level.[24] Because of the household attrition (which deserves closer inspection), I hesitate to emphasize the comparison with the 49.7% figure found for 2008/2009, but I can say that I fail to find support for a decrease in corruption from

[20] For example, the general government final consumption expenditure for China in 2008 was 13.2% of GDP; the same figure for Indonesia was 8.4% (World Bank 2008).
[21] In the 2008 Transparency International Corruption Perceptions Index, which ranks 180 countries in descending order of corruption, China ranks 72 while Indonesia ranks 126.
[22] This simple approach nevertheless would require considerable labor. The IFLS data are separated into numerous files, and variables vary somewhat across years. Also, the best way to aggregate income is not apparent.
[23] The 2014/2015 wave did not record the number of chickens and ducks owned by households, so these controls are absent from the analysis.
[24] The 95% confidence interval is 46.2%-65.6%. Due to its small size, the data do produce statistically significant results for the sample divisions presented in Table 3.

2008 to 2015, despite Indonesia's recent anti-corruption efforts (see Kuris 2012a and 2012b for an overview).

A more fruitful avenue may be to implement more complicated panel SF models that include fixed effects to control better for expenditure preferences. Complicating this approach is the possibility that some households change from the public sector to the private sector and vice versa. Additionally, due to the IFLS's small size, this approach may work better with larger panel data from another source.[25]

References

Aigner, Dennis, CA Knox Lovell, and Peter Schmidt. "Formulation and estimation of stochastic frontier production function models." *Journal of Econometrics* 6.1 (1977): 21-37.

Caudill, Steven B., and Jon M. Ford. "Biases in frontier estimation due to heteroscedasticity." *Economics Letters* 41.1 (1993): 17-20.

Caudill, Steven B., Jon M. Ford, and Daniel M. Gropper. "Frontier estimation and firm-specific inefficiency measures in the presence of heteroscedasticity." *Journal of Business & Economic Statistics* 13.1 (1995): 105-111.

Coelli, Tim. "Estimators and hypothesis tests for a stochastic frontier function: A Monte Carlo analysis." *Journal of productivity analysis* 6.3 (1995): 247-268.

D'Agostino, Ralph, and Egon S. Pearson. "Tests for departure from normality. Empirical results for the distributions of b_2 and $\sqrt{b_1}$." *Biometrika* 60.3 (1973): 613-622.

Donchev, Dilyan, and Gergely Ujhelyi. "What do corruption indices measure?" *Economics & Politics* 26.2 (2014): 309-331.

Freund, Caroline, and Nikola Spatafora. "Remittances, transaction costs, and informality." *Journal of Development Economics* 86.2 (2008): 356-366.

[25] The Indonesian Socioeconomic Survey (SUSENAS) contains a large panel component, but the government charges money for the data and even withholds data. For example, I obtained from the 2008 SUSENAS most of the variables required to conduct this paper's cross-sectional analysis. The government, however, does not publicly offer the household income variable, only the wages earned by individuals. I do not know whether household income is available for other years.

Gorodnichenko, Yuriy, and Klara Sabirianova Peter. “Public sector pay and corruption: Measuring bribery from micro data.” *Journal of Public Economics* 91.5-6 (2007): 963-991.

Hadri, Kaddour. “Estimation of a doubly heteroscedastic stochastic frontier cost function.” *Journal of Business & Economic Statistics* 17.3 (1999): 359-363.

Johnson, Norman L., Samuel Kotz, and N. Balakrishnan. *Continuous univariate distributions, vol. 2.* John Wiley & Sons, 1995.

Jondrow, James, et al. “On the estimation of technical inefficiency in the stochastic frontier production function model.” *Journal of Econometrics* 19.2-3 (1982): 233-238.

Kodde, David A., and Franz C. Palm. “Wald criteria for jointly testing equality and inequality restrictions.” *Econometrica: Journal of the Econometric Society* (1986): 1243-1248.

Kumbhakar, Subal C., Hongren Wang, and Alan P. Horncastle. *A practitioner's guide to stochastic frontier analysis using Stata*. Cambridge University Press, 2015.

Kumbhakar, Subal C., and CA Knox Lovell. *Stochastic frontier analysis*. Cambridge University Press, 2000.

Kuris, Gabriel. “Inviting a Tiger into Your Home: Indonesia Creates an Anti-Corruption Commission with Teeth, 2002-2007.” *Princeton University: Innovations for Successful Societies.* (2012a).

—— “Holding the High Ground With Public Support: Indonesia’s Anti-Corruption Commission Digs In, 2007-2011.” *Princeton University: Innovations for Successful Societies.* (2012b).

Martinez-Bravo, Monica, Priya Mukherjee, and Andreas Stegmann. “The Non-Democratic Roots of Elite Capture: Evidence From Soeharto Mayors in Indonesia.” *Econometrica* 85.6 (2017): 1991-2010.

Martinez-Bravo and Stegmann (Repository of Indonesian Democratic Transition Data). (Accessed on 4 May 2018)

Meeusen, Wim, and Julien van Den Broeck. “Efficiency estimation from Cobb-Douglas production functions with composed error.” *International Economic Review* (1977): 435-444.

OECD, OECD.Stat, (Public Sector, Taxation, and Market Regulation).
doi: https://stats.oecd.org/Index.aspx?DataSetCode=REVIDN
(Accessed on 4 August 2018)

Olken, Benjamin A. “Corruption perceptions vs. corruption reality.” *Journal of Public Economics* 93.7-8 (2009): 950-964.

Nguyen, Quang. “Do natural disasters open a window of opportunity for corruption?” *The Journal of Development Studies* 53.1 (2017): 156-172.

Pissarides, Christopher A., and Guglielmo Weber. “An expenditure-based estimate of Britain's black economy.” *Journal of Public Economics* 39.1 (1989): 17-32.

Ravallion, Martin. “Measuring aggregate welfare in developing countries: How well do national accounts and surveys agree?” *Review of Economics and Statistics* 85.3 (2003): 645-652.

Redwine, Arlo. “Measuring Chinese Corruption With Household Data: A Stochastic Frontier Approach” *Working Paper* (2018).

Saha, Sarani, Poulomi Roy, and Saibal Kar. “Public and private sector jobs, unreported income and consumption gap in India: Evidence from micro-data.” *The North American Journal of Economics and Finance* 29 (2014): 285-300.

Schmidt, Peter, and Tsai-Fen Lin. “Simple tests of alternative specifications in stochastic frontier models.” *Journal of Econometrics* 24.3 (1984): 349-361.

Schneider, Friedrich, Andreas Buehn, and Claudio E. Montenegro. “Shadow economies all over the world: New estimates for 162 countries from 1999 to 2007.” *Handbook on the Shadow Economy* (2011): 9-77.

Shonkwiler, J. Scott, David A. Grigorian, and Tigran A. Melkonyan. “Controlling for the underreporting of remittances.” *Applied Economics* 43.30 (2011): 4817-4826.

Stephens Jr, Melvin. “The consumption response to predictable changes in discretionary income: Evidence from the repayment of vehicle loans.” *The Review of Economics and Statistics* 90.2 (2008): 241-252.

Stevenson, Rodney E. “Likelihood functions for generalized stochastic frontier estimation.” *Journal of Econometrics* 13.1 (1980): 57-66.

United Nations Office on Drugs and Crime (UNODC). *Corruption in the Western Balkans: Bribery as Experienced by the Population.* Vienna: United Nations Office on Drugs and Crime, 2011a.

—— *Corruption in Bosnia: Bribery as Experienced by the Population.* Vienna: United Nations Office on Drugs and Crime, 2011b.

—— *Corruption in Serbia: Bribery as Experienced by the Population.* Vienna: United Nations Office on Drugs and Crime, 2011c.

Wang, Hung-Jen, and Peter Schmidt. “One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels.” *Journal of Productivity Analysis* 18.2 (2002): 129-144.

World Bank, *DataBank,* (World Development Indicators). doi: http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators (Accessed on 8 August 2018)

Zhong, Hai. “Measuring corruption in China: An expenditure-based approach using household survey data.” *Economica* 85.338 (2018): 383-405.

Appendix

Tables and Figures

Table 1: Mean of Variables Used in the Analysis

Variable	Full Sample	Public-Sector Household	Private-Sector Household	*p*-value for Differences
Household income	48.88	69.32	41.29	0.000
Log of household income	17.18	17.77	16.96	0.000
Household expenditure	1.06	1.61	0.86	0.000
Log of household expenditure	13.35	13.80	13.19	0.000
Inside water source	0.66	0.73	0.64	0.000
Outside kitchen	0.30	0.29	0.31	0.427
Household has wood floors	0.45	0.55	0.41	0.000
Household has moderately sized yard	0.63	0.73	0.59	0.000
Household owns a toilet	0.83	0.94	0.79	0.000
Household owns a fridge	0.54	0.74	0.47	0.000
Household owns a stove	0.82	0.88	0.80	0.000
Household owns a television	0.87	0.94	0.84	0.000
No. of chickens owned by household	2.24	4.90	1.25	0.000
No. of ducks owned by household	0.19	0.22	0.17	0.625
Age of household head	38.02	42.39	36.40	0.000
No. of children in household	0.36	0.57	0.29	0.000
Household head has poor health	0.09	0.08	0.10	0.169
Years of education received by household head	9.26	11.72	8.35	0.000
Years of education received by spouse	8.68	10.43	8.03	0.000
Home is owned	0.61	0.69	0.58	0.000
Household pays rent for accommodation	0.13	0.08	0.14	0.000
Household pays education costs	0.64	0.76	0.60	0.000
Household is in an urban area	0.67	0.66	0.68	0.402
Household head is Muslim	0.90	0.87	0.91	0.001
Spouse is working	0.50	0.61	0.46	0.000

Notes: The expenditure and income variables, before logging, are measured in millions of Indonesian rupiah (Rp). The logged values are measured in rupiah. Household expenditure includes only spending on clothing, household supplies, and furniture. Poor health refers to poor self-reported health.

Table 2: ML Estimation Results for the Stochastic Frontier Model

	Coefficient	S.E.
Log of household income	0.258***	0.0217
Assets		
Inside water source	0.120***	0.0419
Outside kitchen	–0.112***	0.0393
Household has wooden floors	0.031	0.0410
Household has moderately sized yard	–0.018	0.0398
Household owns a toilet	0.029	0.0527
Household owns a fridge	0.207***	0.0422
Household owns a stove	0.174***	0.0543
Household owns a television	0.144**	0.0571
No. of chickens owned by household	0.003	0.0024
No. of ducks owned by household	0.014*	0.0077
Demographic controls		
Age of household head	0.006	0.0138
Square of age of household head	–0.000125	0.000162
No. of children in household	0.160***	0.0312
Household head has poor health	0.111*	0.0626
Years of education received by household head	0.021***	0.00643
Years of education received by spouse	0.027***	0.00675
Household controls		
Home is owned	0.108**	0.0456
Household pays rent for accommodation	0.109*	0.0639
Household pays education costs	0.113**	0.0445
Household is in an urban area	–0.158***	0.0467
Household head is Muslim	0.242***	0.0667
Spouse is working	–0.0337	0.0376
Constant	8.124***	0.4351
$\sigma_{u,i}^2$		
Public-sector household	–0.551**	0.2449
Constant	–0.795***	0.2812
$\sigma_{v,i}^2$		
Public-sector household	0.249***	0.0964
Constant	–0.699***	0.0953
N	2,167	

Notes: Dependent variable: logarithm of household expenditure on clothing, household supplies, and furniture. To save space I do not report the results for the district dummies. ***, **, * indicate statistically significant at the 1%, 5% and 10% level, respectively.

Table 3: Estimation Results for the Level of Corruption, by Different Groups and Regions

	Main Results	Shorter Exposure	Longer Exposure	Trust Police	Distrust Police	Poor Districts	Rich Districts
$N^{private}$	1,626	839	937	982	644	590	1,036
N^{public}	541	294	279	403	138	162	379
$\bar{\eta}_i^{private}$	2.077	2.328	0.562	2.823	1.074	2.701	1.404
	(0.020)	(0.038)	(0.002)	(0.041)	(0.011)	(0.055)	(0.012)
$\bar{\eta}_i^{public}$	1.580	1.957	0.000	2.056	0.737	1.988	1.094
	(0.019)	(0.039)	(0.000)	(0.032)	(0.010)	(0.050)	(0.011)
Under-reporting	0.497***	0.371***	0.562***	0.767***	0.338***	0.714***	0.310***
	(0.037)	(0.068)	(0.004)	(0.068)	(0.024)	(0.108)	(0.023)
95% Conf. Interval	0.425-0.570	0.239-0.504	0.554-0.569	0.634-0.900	0.290-0.386	0.502-0.926	0.270-0.350

Note: Standard errors are reported in the parentheses. ***, **, * indicate statistically significant at the 1%, 5% and 10% level, respectively.

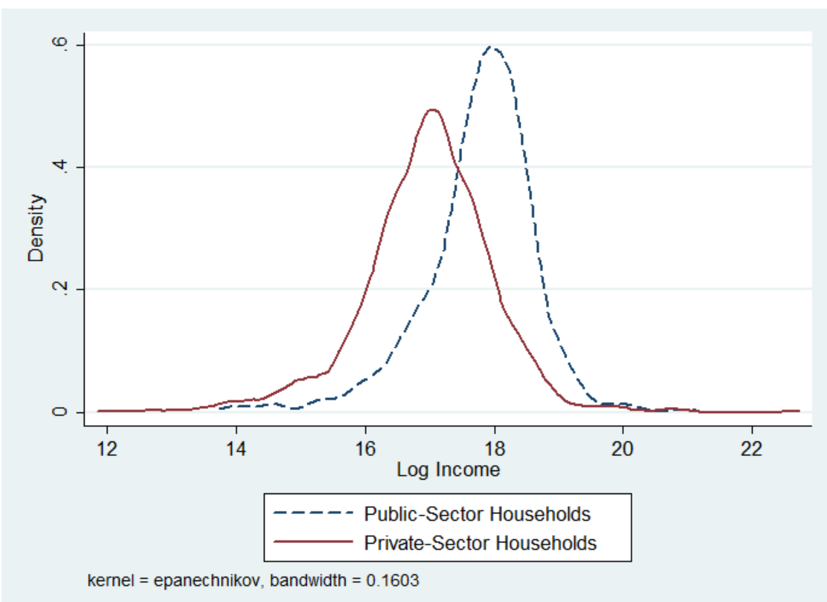


Figure 1: Kernel Density Estimation of the Income Distributions

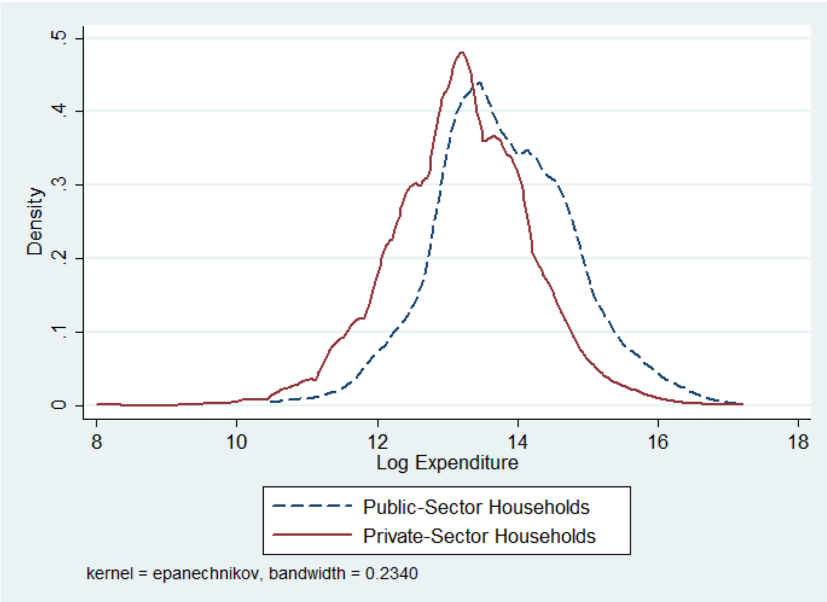


Figure 2: Kernel Density Estimation of the Distributions of the Log of Expenditures on Clothing, Household Items, and Furniture

Appendix: Estimation Issues

Pre-Estimation: Existence of the One-Sided Error

The likelihood function of an SF model is highly nonlinear, and estimation can be difficult. Given this potential difficulty, it is common in the literature to test the validity of the SF specification prior to the ML estimation (Kumbhakar et al. 2015). If support for the SF specification is unfounded, then time is better spent on considering other specifications rather than on the numerical details of the maximization. Schmidt and Lin (1984) propose an OLS residual test to check for the validity of the model's SF specification. The idea behind the test is that for a production-type SF model with composed error $v_i - u_i, u_i \geq 0$ and v_i distributed symmetrically around zero, the residuals from the corresponding OLS estimation should skew to the left. This is true regardless of the distributional function chosen for u_i in the model estimation after the pretesting. A test of the null hypothesis of no skewness as opposed to the alternative hypothesis can thus be constructed using the OLS residuals. If the estimated skewness has the expected sign, rejection of the null hypothesis provides support for the existence of the one-sided error. For the skewness test Schmidt and Lin (1984) suggest a sample-moment-based statistic, commonly referred to as the $\sqrt{b_1}$ test:

$$\sqrt{b_1} = \frac{m_3}{m_2\sqrt{m_2}} \tag{A1}$$

where m_2 and m_2 are the second and third sample moments of the OLS residuals, respectively. The second sample moment of a random variable x is $\sum(x - \bar{x})^2/n$, and the third sample moment is $\sum(x - \bar{x})^3/n$. A result showing $\widehat{\sqrt{b_1}} < 0$ indicates that the OLS residuals are skewed to the left. Under the null hypothesis of no skewness, the statistic should not be statistically different from zero. The distribution of $\sqrt{b_1}$ is nonstandard, and its critical values are tabulated in several studies, e.g., D'Agostino and Pearson (1973).

Coelli (1995) suggests a variant of this test. He notes that under the null hypothesis of no skewness, the third moment of the OLS residuals is asymptotically distributed as a normal random variable with mean 0 and variance $6m_2^3$/N. Thus, the statistic

$$M3T = m_3 \Big/ \sqrt{\frac{6m_2^3}{\mathrm{N}}} \tag{A2}$$

has an asymptotic distribution of a standard normal random variable.

Turning to my data (see main text), I first conduct an OLS estimation of the expenditure function and plot the histogram of the residuals compared to a normal density. The resulting chart is reproduced in Figure A1. There appears to be some evidence of a negative skew, especially in the distribution's central region. To formally examine and test this, I use the skewness statistic. The point estimate of the statistic $\sqrt{b_1}$ (A2) is obtained from the summary statistic of the OLS residuals. The statistic has a value equal to –0.10. The negative sign indicates that the distribution of the residuals skews to the left, which is consistent with a production specification. To assess the statistical significance of the statistic, I conduct the unaltered test as described by D'Agostino, Belanger, and D'Agostino Jr. (1990). The test returns a *p*-value that is 0.054; I reject the null hypothesis of no skewness. Furthermore, the MT3 statistic suggested by Coelli (1995), Equation AB2, equals –1.93. Because it has a normal distribution, the critical value is –1.96, so the result confirms (with a *p*-value of 0.054) the rejection of the null hypothesis of no skewness in the OLS residuals. I have found support for a left-skewed error distribution, and the skewness is statistically significant. I can proceed to the estimation.

Post-Estimation: Existence of the One-Sided Error

Central to the stochastic frontier model is the one-sided error specification, which represents technical inefficiency. It is therefore important to test the existence of the one-sided error for the model. If evidence for the one-sided error specification is not found, the model reduces to a standard regression model for which a simple OLS estimation would suffice. This amounts to a test for the presence of u_i in the model, and a generalized likelihood ratio (LR) test for the null hypothesis of no one-sided error can be constructed based on the log-likelihood values of the OLS (restricted) and the SF (unrestricted) model. Recall that the OLS-residual-based skewness test introduced in the previous section also tests the validity of the one-sided error specification. Although useful as a screening device, the test does not use the information from the distribution functions as the random error. The LR test introduced here is more precise to the specific model I am estimating, but the disadvantage is that it can only be conducted after the ML estimation of the model has been undertaken. The LR test statistic is

$$-2[L(H_0) - L(H_1)], \qquad (A3)$$

where $L(H_0)$ and $L(H_1)$ are log-likelihood values of the restricted model (OLS) and the unrestricted model (SF), respectively, and the degree of freedom equals the number of restrictions in the test. For the half-normal model, the LR test amounts to testing the hypothesis that $\sigma_u^2 = 0$. The complication of the test is that the null hypothesis of $\sigma_u^2 = 0$ is on the boundary of the parameter value's permissible space, and therefore the LR test statistic does not have a standard chi-square distribution. Coelli (1995) shows that, in such cases, the test has a mixture of chi-square distributions. The critical values of the mixed distribution for hypothesis testing are tabulated in Table 1 of Kodde and Palm (1986).

I test for the presence of u_i in the model by constructing the LR test statistic (A3) using the log-likelihood values of the OLS and SF models. The statistic is 14.98, which I compare with a 1% critical value of 9.50. I can reject the null hypothesis of no one-sided error.

The Independence of u_i and v_i

A final estimation-related issue is the independence of u_i and v_i. This assumption is not too restrictive for production models in general because v_i represents shocks outside the control of a firm, and therefore it is unlikely to be related to inefficiency, u_i. One can, however, think of cases in which production risk is captured by the v_i term and risk-taking behavior might be reflected in the inefficiency term. Similarly, one can think of shocks to households that may affect the degree to which they underreport income. For example, households in the public sector may embezzle more funds after local disasters. Approaches exist to handle such nonindependence issues but at the cost of making additional assumptions on the correlation between v_i and u_i. Instead of doing this, I follow most of the production literature and assume independence.

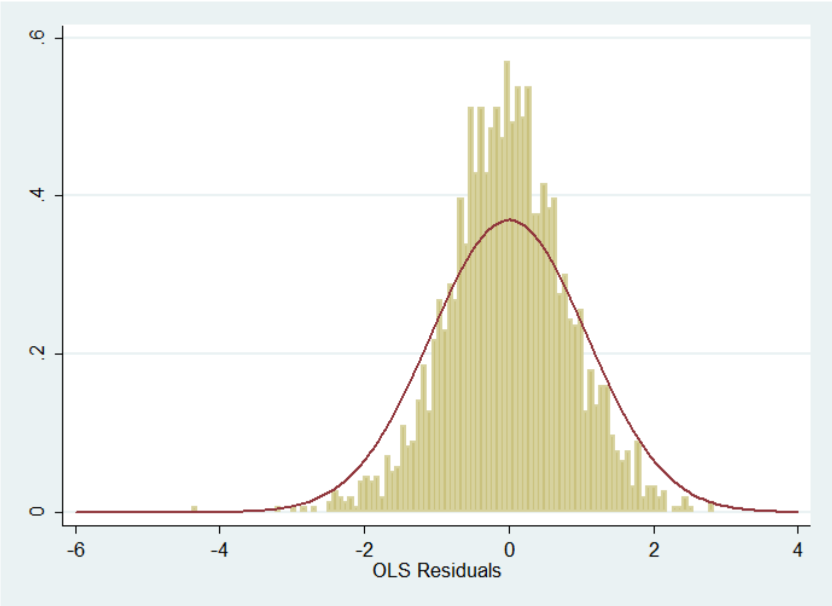


Figure A1: Histogram of OLS Residuals

5 Conclusion

In this dissertation I study corruption and collusion with data derived from a laboratory experiment and household data. In Chapter 1 I test the theory of Compte et al. (2005), who prove the general existence of a collusive equilibrium in which all sellers in a procurement auction bid the government's reserve and submit the highest bribe the agent will accept. In the lab I implement a specific version of their theory to test the extent to which the equilibrium emerges in a simplified setting. Only 13.5% of auctions in the closing periods display the perfectly collusive outcome because only half of subjects attempt such collusion. Yet due to imperfectly collusive bidding, another 45.2% of selling prices are noncompetitive. Therefore in most auctions the selling price is much higher and efficiency much lower than would be expected if subjects simply raised their honest-auction bids by the maximum bribe. Much of Chapter 1 tries to explain why many subjects, having chosen to bid above a certain amount, fail to realize that the most sensible bid is the reserve. Confusion and gender seem to be the main reasons, though some subjects learn from others. My essay, while contributing to the corruption literature, holds a firmer place in the experimental literature on collusion, in which little evidence exists of tacit collusion among sellers who are continually rematched. An important difference between my study and previous auction studies is that bounded rationality, much more than trust, is the limiting factor for collusion.

My existence proof of Compte et al.'s collusive equilibrium in the lab does not determine whether the theory is applicable to more complicated markets. Even so, I argue that real-world firms have a greater chance at optimal collusion. Especially in the world of lucrative government contracts, corrupt firms probably often do wish to pay a bribe beyond what the agent will accept, and this market imperfection probably does often lead to collusion, even if such collusion better resembles one of the more complicated models described by Compte et al. Furthermore, firms are less likely than lab subjects to face hindrances such as confusion and joy of winning (which are not accounted for in Compte et al.'s model anyhow), and thus may achieve collusion more easily. Firms also may choose explicit collusion over tacit collusion due to the latter's fragility. One possibility for future research is allowing subjects to communicate with each other, which would likely lead to less confusion and more perfect collusion, albeit explicit.

My inspiration for Chapter 2 is Zhong (2018), who claims to provide the first objective estimate of the extent of corruption in China (in 2002). Using a household survey, Zhong first uses OLS to estimate an expenditure function for households without opportunities for corruption, then inverts the function to forecast the true income of households with opportunities. The difference between true income and reported income, he assumes, is the value of illegal income. According to his preferred measure of corruption, based on clothing expenditure, the true incomes of households with more opportunities for corruption are, on average, about 20% higher than their reported income. This estimate, however, is based on a quadratic expenditure function requiring a method that, at least according to my analysis in Chapter 2's Appendix A, seems to be highly sensitive to sample selection. On the other hand, I easily replicate Zhong's estimate based on expenditure of household equipment and services, which does not require a quadratic term. That estimate indicates about 15% instead of 20%, although Zhong cannot say whether the estimates are statistically significant. Using the same household survey, I apply Stochastic Frontier (SF) analysis to produce a statistically testable estimate corresponding to Zhong's 15% measurement. The SF method indicates about 10%, which is significant at the 1% level. Regarding the difference in our point estimates, I argue that my method alone may partially correct for bias in a key parameter — and thus may be closer to the truth. Even so, this paper primarily vindicates Zhong's approach to measuring corruption because most of our robustness checks follow similar patterns.

Clearly, a clean benchmark group is hard to find. These estimations, therefore, are best thought of either as lower-bound estimations or as relative measures of corruption (my preference). The value of the methods is their ease and affordability in comparison with micro-level analysis based on administrative records, special-purpose surveys, or field experiments. With just the variables found in many household surveys, researchers can explore specific issues such as the relative corruption between specific groups or regions, or the correlation between a community's concern for corruption and the extent of corruption. Here, for example, my estimates seem to indicate a negative correlation. Researchers also can use the SF method to measure tax evasion. Policymakers, when allocating funds for fighting corruption, could benefit from objective measures of corruption's intensity, perhaps by region.

In Chapter 3 I apply the method presented in Chapter 2 to an Indonesian dataset. I find that the true incomes of public-sector households are, on average, about 50% higher than their reported income, providing a rare measure of corruption's magnitude. I then divide the sample to support the findings of Martinez-Bravo et al. (2017), who exploit the fact that district mayors of the Suharto regime could finish their terms during the democratic transition. Suharto appointed these mayors in 1994, 1995, 1996, or 1997, leading to exogenous variation in corruption exposure. When I restrict my sample to shorter-exposed districts (those with mayors whose appointments were in 1996 or 1997), my corruption measurement falls to 37.1%; when I restrict my sample to longer-exposed districts, my measurement rises to 56.2%. My results provide a correlation absent from Martinez-Bravo et al. (2017): that between mayor exposure and corruption magnitude.

Regarding future research, the IFLS is a longitudinal study, so possibly I could employ the SF method to study corruption estimates across time, using the years 1993/1994, 1997, 2000, 2008/2009 (the data of chapter 4), and 2014/2015. A simple approach is to compare cross-sectional measurements. A more fruitful avenue may be to implement more complicated panel SF models that include fixed effects to control better for expenditure preferences. Complicating this approach is the possibility that some households change from the public sector to the private sector and vice versa. Additionally, due to the IFLS's small size, this approach may work better with larger panel data from another source.

Appendix

UNIVERSITY OF ARKANSAS

Office of Research Compliance
Institutional Review Board

December 2, 2015

MEMORANDUM

TO:	Arlo Redwine Peter McGee
FROM:	Ro Windwalker IRB Coordinator
RE:	New Protocol Approval
IRB Protocol #:	15-11-364
Protocol Title:	*Corruption and Tacit Collusion in Procurement Auctions*
Review Type:	☒ EXEMPT ☐ EXPEDITED ☐ FULL IRB
Approved Project Period:	Start Date: 12/02/2015 Expiration Date: 12/01/2016

Your protocol has been approved by the IRB. Protocols are approved for a maximum period of one year. If you wish to continue the project past the approved project period (see above), you must submit a request, using the form *Continuing Review for IRB Approved Projects*, prior to the expiration date. This form is available from the IRB Coordinator or on the Research Compliance website (https://vpred.uark.edu/units/rscp/index.php). As a courtesy, you will be sent a reminder two months in advance of that date. However, failure to receive a reminder does not negate your obligation to make the request in sufficient time for review and approval. Federal regulations prohibit retroactive approval of continuation. Failure to receive approval to continue the project prior to the expiration date will result in Termination of the protocol approval. The IRB Coordinator can give you guidance on submission times.

This protocol has been approved for 108 participants. If you wish to make *any* modifications in the approved protocol, including enrolling more than this number, you must seek approval *prior to* implementing those changes. All modifications should be requested in writing (email is acceptable) and must provide sufficient detail to assess the impact of the change.

If you have questions or need any assistance from the IRB, please contact me at 109 MLKG Building, 5-2208, or irb@uark.edu.

109 MLKG • 1 University of Arkansas • Fayetteville, AR 72701-1201 • (479) 575-2208 • Fax (479) 575-6527 • Email irb@uark.edu
The University of Arkansas is an equal opportunity/affirmative action institution.