Paying Over the Odds at the End of the Fiscal Year. Evidence from Ukraine

Online Appendices

Margaryta Klymak Stuart Baumann

November 26, 2024

Abstract

This document contains the appendices for "Paying Over the Odds at the End of the Fiscal Year. Evidence from Ukraine" which is published in the Journal of Econometrics.

The paper can be found (with open access) at:

https://doi.org/10.1016/j.jeconom.2024.105903

Appendices

A Derivation of Linear Mappings in Second Step of Neural Network Procedure

The demand equation is as per equation 6:

$$D(p) = n_{3,l,b} + n_{2,l,b} \exp(-n_{1,l,b} \times \frac{(p_{l,b} - \tilde{p}_{l,b})}{v_l}) \tag{A.1}$$

Now we can derive an expression for the implied cost from a price using equation 5:

$$c = p_{l,b} - v_l \left[\frac{1}{n_{1,l,b}} + \left(\frac{n_{3,l,b}}{n_{2,l,b}} \right) \frac{1}{n_{1,l,b} \exp(-n_{1,l,b} \times \frac{(p_{l,b} - \tilde{p}_{l,b})}{v_l})} \right] \tag{A.2}$$

We can take the derivative of equation A.2 with respect to $p_{l,b}$ to get the marginal cost as a function of price:

$$\frac{\partial c}{\partial p_{l,b}} = 1 - \left(\frac{n_{3,l,b}}{n_{2,l,b}}\right) \frac{1}{\exp(-n_{1,l,b} \times \frac{(p_{l,b}-\tilde{p}_{l,b})}{v_l})} \tag{A.3}$$

To accord with economic intuition we require that this is positive in the range of $p_{l,b} - \tilde{p}_{l,b}$ values that are seen in our dataset. This condition is

$$\left(\frac{n_{3,l,b}}{n_{2,l,b}}\right) \leq \exp(-n_{1,l,b} \times \frac{(p_{l,b} - \tilde{p}_{l,b})}{v_l}) \tag{A.4}$$

We choose to use $\frac{(p_{l,b}-\tilde{p}_{l,b})}{v_l} = \nabla = 0.25$ for the purposes of this restriction as this covers 97.5% of our data.

$$\left(\frac{n_{3,l,b}}{n_{2,l,b}}\right) \leq \exp(-n_{1,l,b} \times \nabla) \tag{A.5}$$

Now returning to equation A.2, we will divide by price so that we get cost as a fraction of the final quoted price:

$$\frac{c}{p_{l,b}} = 1 - \frac{v_l}{p_{l,b}} \left[\frac{1}{n_{1,l,b}} - \left(\frac{n_{3,l,b}}{n_{2,l,b}}\right) \frac{1}{n_{1,l,b} \exp(-n_{1,l,b} \times \frac{(p_{l,b}-\tilde{p}_{l,b})}{v_l})}\right] \tag{A.6}$$

We want to impose that the gradients of the demand function must be such that the costs as a fraction of the same price are somewhere between $0 \leq L < 1$ and 1. The condition for this is:

$$\exp(-n_{1,l,b} \times \frac{(p_{l,b} - \tilde{p}_{l,b})}{v_l}) \left[\frac{p_{l,b}}{v_l} n_{1,l,b}(1 - L) - 1\right] - \left(\frac{n_{3,l,b}}{n_{2,l,b}}\right) \geq 0 \tag{A.7}$$

Clearly from equation A.6 this is not possible for all prices as prices near zero will lead to this condition not being met. We will find restrictions such that costs are within this range for all prices $\frac{p}{v_l} \geq p_*$. A necessary condition for equation A.7 is that the result of the expression in the square brackets is positive. This implies a condition for $n_{1,l,b}$:

$$n_{1,l,b} \geq \frac{1}{(1 - L)p_*} \tag{A.8}$$

There are no pressing restrictions to place on $n_{2,l,b}$ other than its positivity (required so that higher prices lead to lower demand). We do have a restriction for $\left(\frac{n_{3,l,b}:}{n_{2,l,b}}\right)$ however

$$\left(\frac{n_{3,l,b}}{n_{2,l,b}}\right) \leq \exp(-n_{1,l,b} \times \frac{(p_{l,b} - \tilde{p}_{l,b})}{v_l}) \left[\frac{p_{l,b}}{v_l} n_{1,l,b}(1 - L) - 1\right] \tag{A.9}$$

Putting these restrictions together we use the following expressions to map the outputs of the neural network to the n parameters:

$$n_{1,l,b} = \frac{1}{p_*(1 - L)} + \frac{2}{p_*} \times o_1 \tag{A.10}$$

Additionally, we require the demand curve to be downward sloping and hence $n_{2,l,b}$ is positive. We allow it to go to the range [0.01,4.01] where the lower limit is chosen to ensure positivity while the upper limit is sufficiently high that it does not often bind.

$$n_{2,l,b} = 0.01 + 4 \times o_2 \tag{A.11}$$

In terms of $n_{3,l,b}$ there are two constraints in equations A.5 and A.9. As a result, we use the following mapping for $n_{3,l,b}$.

$$n_{3,l,b} = n_{2,l,b} \left[-0.5 + o_1 \left(0.5 + \min\left(\exp(-n_{1,l,b} \times \nabla)\ ,\ \exp(-n_{1,l,b} \times \nabla)\left[p_* n_{1,l,b}(1 - L) - 1\right]\right)\right)\right] \tag{A.12}$$

B External Validity

We examine whether our main findings of paying over the odds hold in other countries. We replicate our benchmark specification for Japan where the fiscal year ends on the 31st of March and for Georgia (another post-Soviet country) where the fiscal year overlaps with that of Ukraine.

B.1 Japan

First, we start with the Japanese government's procurement data from Japan which spans January 2007 through to December 2021. The departments of the Japanese government most frequently rely on an open tendering procedure with nearly 80% of goods and services procured via this procedure. Our dataset consists of bids submitted by construction firms participating in auctions by the Ministry of Land, Infrastructure and Transportation in Japan and the sample consists of roughly 867,000 procurement auctions. Investing in the construction sector constitutes almost 20% of Japan's GDP and alone public investment comprised 45.6% of the overall construction investment (Nakabayashi, 2013). The auction format used in our data is a first-price sealed-bid auction with a hidden reserve price (Kawai and Nakabayashi, 2022). The dataset comprises details about the reserve price, all bids, bidders identity, auction date, location and project type. This procedure involves publishing a tender notice that is publicly disclosed, allowing eligible suppliers (suppliers that meet various qualifications to bid on government work) to participate in the bidding process.

We present summary statistics in Table B.1[1] which depicts a similar pattern we observe for Ukraine: the number of tenders and total aggregate expenditure is the highest in March, the last month of the fiscal year. The price ratio is also highest in this month as well. While Japan has a very different institutional framework and more competitive auctions, we still observe lower-value lots and higher price ratios at their fiscal year-end. This implies a degree of external validity of our findings. We then estimate the following specification which is very similar to our baseline specification equation 3 in the main body of the paper:[2]

$$\text{Price ratio}_{ifdy} = \beta_0 + \gamma_1 \text{Last month}_{ifdy} + \phi_d + \eta_f + \psi_y + \epsilon_{ifdy} \tag{B.1}$$

where the dependent variable Price ratio_{ifdy} is defined as per equation 4, reproduced below:

$$\text{Price ratio}_{ifdy} = \left[\frac{\text{Winning price}_{ifdy}}{\text{Reserve price}_{ifdy}}\right] \times 100 \tag{B.2}$$

As in the main body of the paper, our regressor of interest is Last month_{ifdy} which is a dummy variable assuming a value of one in instances where a purchase occurred in March and zero otherwise. Additionally, we incorporate the purchasing department d, the supplier winning in the bidding process f, and the year y fixed effects. The term ϵ_{ifdy} represents the statistical error component in the model. The results of this specification can be found in Table B.2. Our findings for Japan further confirm earlier results for Ukraine and show the increased year-end spending.

B.2 Georgia

Georgia's public procurement contracts correspond to approximately 10% of GDP (World Bank Group, 2018) and have historically been criticised for the lack of transparency, high risk of

[1] Note that the average job size in Table B.1 is higher than for Ukraine or Georgia as our Japanese data considers construction projects.

[2] Indeed the only difference between these specifications is that we do not control for industry fixed effects in Japan given that all procurements are in the construction industry.

Table (B.1) Summary Statistics for Japan's Central Government Procurement 2007 – 2021

Month	Number of Bidders	Number of Lots	Total Notional Won	Total Notional Reserve	Mean Price Ratio
January	7.05	5,433	1,597,000	1,777,000	90.79
February	6.66	14,718	2,927,000	3,249,000	91.17
March	5.55	33,952	4,684,000	5,088,000	92.11
April	5.47	8,184	936,300	1,026,000	91.13
May	6.29	6,576	963,900	1,063,000	91.24
June	6.65	10,885	1,609,000	1,773,000	91.34
July	6.46	13,398	2,002,000	2,195,000	91.50
August	6.38	13,253	1,968,000	2,164,000	91.62
September	5.92	16,537	2,187,000	2,405,000	91.49
October	5.39	9,019	1,224,000	1,346,000	91.68
November	5.50	6,128	1,009,000	1,049,000	91.36
December	6.24	5,103	1,013,000	1,118,000	91.13

Note: "Mean number of bidders" represents the average number of participants per tender in a given month. "Mean number lots won" reports the average number of lots that were successfully completed per year. "Total notional won" reports the award amount (in millions of ¥) while "Total notional reserve" is the reservation amount (in millions of ¥) per month. The price ratio is defined as a winning price for a lot from each procurement participant over the reservation value of that lot.

Table (B.2) The Year End Effect for Japan

	Ratio			
	(1)	(2)	(3)	(4)
Last Month	0.476***	0.525***	0.416***	0.133**
	(0.0821)	(0.0796)	(0.0690)	(0.0544)
Year FE	No	Yes	Yes	Yes
Buyer FE	No	No	Yes	Yes
Firm FE	No	No	No	Yes
Industry FE	143149	143141	143141	139742
N	0.0694	0.0588	0.152	0.353

Note. The sample is composed of winners only. The dependent variable in all columns is the price ratio. Last month is a dummy variable that takes the value one if the tender was finalised in March and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects and firm FE are firm fixed effects. All coefficients should be interpreted as percentage changes to the price (relative to the expected price). Robust standard errors are adjusted at the buyer level. Coefficients that are significantly different from zero are denoted by the following system: $* \ p < 0.10$; $** \ p < 0.05$; $*** \ p < 0.01$.

corruption and restricted competition. To address these issues, the State Procurement Agency of Georgia has developed and introduced a mandatory e-procurement system which replaced the previous paper based approach in December 2010. The digitisation of processes has heightened competition among suppliers, diminished bureaucratic hurdles, and provided equal access to tenders (World Bank Group, 2018). Similar to Ukraine, Georgia's system also operates under motto "everyone sees everything" where public has access to tenders in real time.

The new procurement system has been recognised as a success by various international organisations, including the United Nations, the European Bank for Reconstruction and Development, the OECD, and The World Bank (State Procurement Agency of Georgia, 2015). According to Transparency International (2012), Georgia improved on the 2012 Corruption Perception Index in great part due to *"the introduction of an innovative electronic procurement system"*.

Table (B.3) Summary Statistics for Georgia's Government Main Procurement 2011 – 2022

Month	Number of Bidders	Number of Lots	Total Notional Won	Total Notional Reserve	Mean Price Ratio
January	2.11	23,783	2,625,476	2,592,541	90.57
February	2.17	26,695	3,027,598	3,191,740	90.10
March	2.14	23,790	2,726,181	2,990,786	89.70
April	2.12	20,234	2,394,381	2,689,874	89.91
May	2.05	19,096	2,223,125	2,451,293	90.23
June	2.04	20,270	2,600,552	2,939,590	88.95
July	1.98	21,332	2,669,176	2,948,385	90.75
August	1.91	21,598	2,787,007	3,044,713	91.16
September	1.94	19,347	2,624,059	2,855,136	91.23
October	2.00	17,530	2,210,598	2,460,459	89.67
November	2.03	18,426	3,106,397	3,300,363	91.87
December	1.91	32,249	4,724,968	4,565,184	92.06

Note: "Mean number of bidders" represents the average number of participants per tender in a given month. "Number lots won" reports the average number of lots that were successfully completed in that month. "Total notional won" reports the award amount (in thousands of Georgian Lari) while "Total notional reserve" is the reservation amount (in thousands of Georgian Lari) per month. The price ratio is defined as a winning price for a lot from each procurement participant over the reservation value of that lot weighted by the procurement size.

There are a number of procedures followed by Georgia in its public procurement depending on the anticipated cost and the requirements of the procuring agency. On one side there are simplified procedures for relatively simple procurements with direct purchases being possible for procurements up to a value of 10,000 lari (or about \$3,700). Larger procurements up to 150,000 lari (or about \$ 56,000) should be done under a 3-stage reverse electronic auction with a minimum term of 7 calendar days. Above that there are first price auctions with the exact specifications and time to advertise the tender being dependent on the specifics of the auction.

Table (B.4) The Year End Effect for Georgia

	Ratio				
	(1)	(2)	(3)	(4)	(5)
Panel A: Entire sample					
Last month	0.655***	0.661***	0.546***	0.243**	0.195*
	(0.176)	(0.176)	(0.150)	(0.108)	(0.106)
Observations	253,363	253,363	253,363	253,363	253,358
R^2	0.0002	0.006	0.054	0.369	0.379
Panel B: Auctions only					
Last month	0.937***	0.987***	0.945***	0.616***	0.538***
	(0.234)	(0.236)	(0.205)	(0.144)	(0.141)
Year FE	No	Yes	Yes	Yes	Yes
Buyer FE	No	No	Yes	Yes	Yes
Seller FE	No	No	No	Yes	Yes
Industry FE	No	No	No	No	Yes
Observations	154,134	154,134	154,134	154,134	154,134
R^2	0.0003	0.005	0.057	0.363	0.374

Note. The sample is composed of winners only. The dependent variable in all columns is the price ratio. Last month is a dummy variable that takes the value one if the tender was finalised in December and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are sectoral fixed effects. All coefficients should be interpreted as percentage changes to the price (relative to the expected price). Standard errors are in parentheses and are clustered by government departments. Coefficients that are significantly different from zero are denoted by the following system: $* \, p < 0.10$; $** \, p < 0.05$; $* * * \, p < 0.01$.

We focus on procurement tenders for the January 2011 – December 2022 period obtained from the State Procurement Agency. The dataset captured information on over 259 thousand tenders where we observed the expected value, the winning value, the content of purchase, the identity of the buyer, as well as the bids made by tender participants. Expected price is disclosed before the auction and bidders are allowed to price their bids below or above specified price. Crucially, we also observe tender announcement and auction dates which allows us to identify whether the procedure took place at the end of the fiscal year in Georgia (the same as in Ukraine) in December. Table B.3 provides the summary statistics of the key indicators for the entire sample in the top panel and for competitive auctions in the bottom panel.

We estimate the empirical specification of equation B.3 reproduced below which mirrors our baseline specification in equation 3:

$$\text{Price ratio}_{ifdys} = \beta_0 + \alpha_1 \text{Last month}_{ifdys} + \phi_d + \eta_f + \psi_y + \eta_s + \epsilon_{ifdys} \tag{B.3}$$

where Price ratio$_{ifdys}$ is the dependent variable defined as per equation 4, reproduced below:

$$\text{Price ratio}_{ifdys} = \left[\frac{\text{Winning price}_{ifdys}}{\text{Expected price}_{idys}}\right] \times 100 \tag{B.4}$$

Our regressor of interest is Last month$_{ifdys}$ which is dummy variable taking a value of one if a purchase took place in December and zero otherwise. We also control for the buying department d, the bidding supplier f, the year y and industry s, while ϵ_{ifdys} is the statistical error term.

We present the results in Table B.4 where we gradually increase fixed effects before we arrive at column (4) which corresponds to specification B.3. Two main conclusions can be drawn from these findings: (1) paying over the odds at the fiscal year-end is also present in Georgia; (2) the magnitude of the effect is very similar to that found in Ukraine.

C Additional Extensions

C.1 Homogeneous Goods

We might expect suppliers bidding on selling homogenous goods to exhibit different behaviour at year-end as the lack of differentiation means markets for these goods are closer to perfect competition than monopolistic competition. In extreme cases, the presence of liquid commodity markets in some good areas may reduce or eliminate the year-end price change effect (as we observed for the relatively homogenous petroleum products, fuel, and energy product categories in Figure 4). For example, Bandiera, Prat and Valletti (2009) consider the Italian procurement system and find heterogeneity in prices paid by government agencies for homogenous goods. These differences in prices were mostly due to "passive waste", where buyers do not make full use of a centralised purchasing system in which favourable prices are pre-negotiated with suppliers.

We classified the good codes in our sample based on their level of heterogeneity. We assign a good code as homogeneous when there is relatively little vertical or horizontal heterogeneity within the good category. Some examples of homogenous goods are "printer paper", "rice", "golf balls" and "cement", while "crustaceans", "trucks" and "golf clubs" are classified as being heterogeneous. This method of categorisation is not perfect but it is sufficient such that the good we identify as homogeneous is more homogeneous than other goods. Due to the inability to perfectly separate homogeneous from heterogeneous goods, however, we draw attention to the difference between the two interaction coefficients in the regressions rather than their absolute levels.

We then interact the homogeneous and heterogeneous categorisation of a good category with last month dummy and present the results in Table C.1. The results indicate that while the aforementioned strong last month effect is apparent for heterogeneous good categories, this rise in prices at the end of the year is much lower for homogeneous goods.

Table (C.1) The Year-End Effect for Homogeneous and Heterogeneous Goods

	Dependent variable:				
	Price Ratio				
	(1)	(2)	(3)	(4)	(5)
Last Month × Heterogeneous	1.125***	0.849***	0.694***	0.634***	0.626***
	(0.022)	(0.014)	(0.012)	(0.011)	(0.011)
Last Month × Homogeneous	0.522***	0.375***	0.315***	0.300***	0.301***
	(0.029)	(0.025)	(0.019)	(0.019)	(0.019)
Homogeneous	−0.206***	−0.411***	0.125***	0.088***	
	(0.023)	(0.015)	(0.010)	(0.009)	
Observations	11,967,900	11,967,900	11,967,900	11,967,900	11,967,900
R^2	0.026	0.143	0.291	0.330	0.333
Year FE	Yes	Yes	Yes	Yes	Yes
Buyer FE	No	Yes	No	Yes	Yes
Firm FE	No	No	Yes	Yes	Yes
Industry FE	No	No	No	No	Yes

Note: The sample is composed of winners of competitive and non competitive lots. The dependent variable in all columns is the price ratio as defined in equation 4. Last month is a dummy variable that takes the value one if the tender was finalised in December and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are eight-digit sectoral fixed effects. The specification defined in equation 2 is presented in the last column. All coefficients should be interpreted as percentage changes to the price (relative to the expected price). Standard errors are in parentheses and are clustered by government departments. *p<0.1; **p<0.05; ***p<0.01.

C.2 Contract Renegotiation

Given the sharp increase in the number of tenders being advertised at the end of the fiscal year, it is possible that some of these tenders have been rushed. For example, Liebman and Mahoney (2017) suggested this as a reason why the end-of-fiscal-year spending might have been of lower quality than other spending.

We investigate this possibility using two variables that are likely to be related to rushing. First, the final price paid by the government could have changed after the contract was awarded. Second, a contract could be renegotiated after its award, which may or may not be accompanied by a change in price. We estimate the percentage price change between awarded and the final price, a flattened variable indicating whether a price has changed (+1 for a price increase, 0 for no change and −1 for a decrease), and whether or not a contract was renegotiated in the three panels of Table C.2. These regressions are all estimated on competitive lots as we only have this information for these tenders, and we do condition on the number of bidders.

Table (C.2) The Year-End Effect for Contract Renegotiation

	Dependent variable (DV):				
	(1)	(2)	(3)	(4)	(5)
Panel A: Percentage price change DV					
Last Month	-0.007^{***}	-0.004^{***}	-0.002^{***}	-0.001^{**}	-0.001^{**}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	874,357	874,357	874,357	874,357	874,357
R^2	0.010	0.126	0.156	0.229	0.234
Panel B: Binary price change DV					
Last Month	-0.021^{***}	-0.013^{***}	-0.010^{***}	-0.005^{**}	-0.005^{**}
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Observations	874,357	874,357	874,357	874,357	874,357
R^2	0.020	0.142	0.175	0.247	0.253
Panel C: Contract renegotiation DV					
Last Month	0.018^{***}	0.008^{***}	-0.008^{***}	-0.013^{***}	-0.011^{***}
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Observations	874,357	874,357	874,357	874,357	874,357
R^2	0.340	0.416	0.477	0.519	0.524
Year FE	Yes	Yes	Yes	Yes	Yes
Buyer FE	No	Yes	No	Yes	Yes
Firm FE	No	No	Yes	Yes	Yes
Industry FE	No	No	No	No	Yes

Note: The sample is composed of winners of competitive and non-competitive lots. The dependent variable in Panel A is the percentage price change between winning and contract value; in Panel B is a dummy variable taking value one if the price changed and zero otherwise; in Panel C the dependent variable is whether the contract was renegotiated. Last month is a dummy variable that takes the value 1 if the tender was finalised in December and 0 otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are eight-digit sectoral fixed effects. The specification defined in equation 2 is presented in the last column. All coefficients should be interpreted as percentage changes to the price (relative to the expected price). Standard errors are in parentheses and are clustered by government departments. *p<0.1; **p<0.05; ***p<0.01.

The results reveal that price decreases are more likely at the end of the fiscal year, which could be a result of elevated prices at the time of the advertisement. We also cannot find any robust evidence for a different likelihood of renegotiation for contracts awarded at the end of the fiscal year.

C.3 New Suppliers

A plausible mechanism for increased prices is through established suppliers reaching their capacity and government departments needing to employ new and potentially less cost efficient suppliers. For example, Hyndman, Jones and Pendlebury (2007) reported that UK budget managers were not able to use lower-cost suppliers as they were too busy at the end of the fiscal year. As a result, they switched to new suppliers that were more expensive. We first define a "new firm" as one that a government department has not awarded a job to in the preceding 365 days. Otherwise, suppliers are denoted as "old suppliers" as there was a business relationship within the last year. We then test this hypothesis in two steps. First, we check whether new suppliers are more likely to win a tender at the year-end. Second, we compare prices charged by new and existing suppliers. We interact the last month dummy variable with new suppliers and old suppliers and present the results in Table C.3.

Table (C.3) The Year-End Effect for New and Old Firms

	Dependent variable (DV):				
	(1)	(2)	(3)	(4)	(5)
Panel A: New firm winner DV					
Last Month	0.061***	0.043***	0.039***	0.028***	0.028***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	12,003,212	12,003,212	12,003,212	12,003,212	12,003,212
R^2	0.050	0.229	0.340	0.460	0.463
Panel B: Price ratio DV					
New Firm × Last Month	0.685***	0.641***	0.443***	0.455***	0.458***
	(0.021)	(0.018)	(0.015)	(0.014)	(0.015)
Last Month	0.798***	0.561***	0.487***	0.426***	0.419***
	(0.022)	(0.014)	(0.012)	(0.011)	(0.011)
New Firm	−1.198***	−1.017***	−0.755***	−0.814***	−0.813***
	(0.021)	(0.014)	(0.013)	(0.010)	(0.010)
Observations	11,967,900	11,967,900	11,967,900	11,967,900	11,967,900
R^2	0.030	0.145	0.292	0.331	0.334
Year FE	Yes	Yes	Yes	Yes	Yes
Buyer FE	No	Yes	No	Yes	Yes
Firm FE	No	No	Yes	Yes	Yes
Industry FE	No	No	No	No	Yes

Note: The sample is composed of winners of competitive and non-competitive lots. The dependent variable in Panel A is a dummy taking a value of one if a new firm won tender and zero otherwise. A new firm is defined as one that a government department has not awarded a job to in the preceding 365 days. The dependent variable in all columns in Panel B is the price ratio as defined in equation 4. Last month is a dummy variable that takes the value one if the tender was finalised in December and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are eight-digit sectoral fixed effects. The specification defined in equation 2 is presented in the last column. All coefficients should be interpreted as percentage changes to the price (relative to the expected price). Standard errors are in parentheses and are clustered by government departments. *p<0.1; **p<0.05; ***p<0.01.

In short, we cannot find any support for Hyndman, Jones and Pendlebury's (2007) survey findings in our sample of Ukrainian public procurement lots. We do find that new suppliers are more likely to win a tender at year-end and that they increase their prices more at year-end than other suppliers. Even with their larger year-end price increase however new suppliers remain cheaper than existing suppliers at year-end. This may reflect that given existing suppliers are busy at year-end, new suppliers do not need to cut prices as much to sell to the department.

C.4 Heterogeneity by Procurement Size

To further explore heterogeneity in terms of differential year-end price increases according to lot size, we present a number of findings in Table C.4. We split the year-end dummy by the size of the lot. The buckets were set into below 50K for small value lots, between 50-200K lots (where government agencies are mandated to purchase through Prozorro), between 200K and 500K lots (where a heightened level of competition is required. It can be seen that for all subsets of lots and across fixed effects specifications, the prices are higher at the end of the fiscal year.

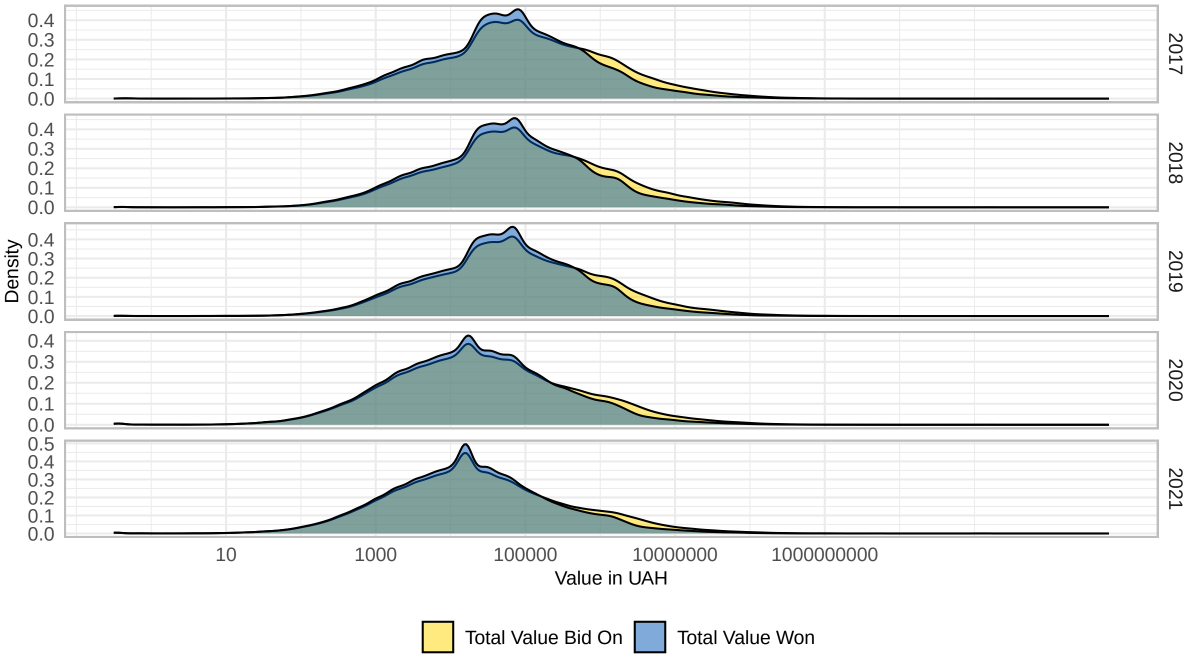
Table (C.4) The Year-End Effect by Lot Size

	Dependent variable:				
	Price Ratio				
	(1)	(2)	(3)	(4)	(5)
Below 50K	4.290***	3.482***	4.221***	3.951***	4.199***
	(0.111)	(0.079)	(0.070)	(0.061)	(0.062)
Between 200K and 500K	0.254***	0.114*	0.637***	0.594***	0.675***
	(0.091)	(0.062)	(0.052)	(0.047)	(0.046)
Between 50K and 200K	1.978***	1.392***	2.167***	1.992***	2.155***
	(0.103)	(0.075)	(0.060)	(0.055)	(0.055)
Last month below 50K	0.693***	0.531***	0.425***	0.391***	0.385***
	(0.017)	(0.012)	(0.010)	(0.009)	(0.009)
Last month 50K-200K	1.677***	1.419***	1.109***	1.030***	1.021***
	(0.057)	(0.047)	(0.039)	(0.037)	(0.037)
Last month 200K-500K	2.180***	1.881***	1.331***	1.278***	1.257***
	(0.096)	(0.090)	(0.073)	(0.073)	(0.072)
Last month 500K and above	1.059***	0.894***	0.719***	0.724***	0.721***
	(0.099)	(0.095)	(0.070)	(0.070)	(0.069)
Observations	11,967,900	11,967,900	11,967,900	11,967,900	11,967,900
R^2	0.054	0.158	0.305	0.341	0.344
Year FE	Yes	Yes	No	Yes	Yes
Buyer FE	No	Yes	No	Yes	Yes
Firm FE	No	No	Yes	Yes	Yes
Industry FE	No	No	No	No	Yes

Note: The sample is composed of winners of competitive and non-competitive lots. The dependent variable in all columns is the price ratio as defined in equation 4. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are eight-digit sectoral fixed effects. All coefficients should be interpreted as percentage changes to the price (relative to the expected price). Standard errors are in parentheses and are clustered by government departments. *p<0.1; **p<0.05; ***p<0.01.

We also plotted the kernel density of the expected values of lots by year. Figure C.1 shows that the center of the distribution has moved to the left in 2020 as it became compulsory to report all low value transactions.

Figure (C.1) Density of Firms' Winning and Bidding Amounts



Note: This figure shows the size of lots being awarded on Prozorro between 2017 and 2021. These figures are expressed in Hryvnia and are deflated by Ukrainian CPI. It can be noted that the center of the distribution has shifted downward which reflects the use of the Prozorro system for an increasing fraction of procurement lots.

C.5 Alternative Empirical Approach

We present an alternative empirical approach that examines whether government departments pay over the odds at the year-end. The specification we use is as follows:

$$\begin{aligned}\text{Log bid price}_{ifdys} = \beta_0 + \beta_1 \text{Last month}_{ifdys} + \beta_2 \text{Log expected price}_{ifdys} + \\ + \phi_d + \eta_f + \psi_y + \eta_s + u_{ifdys}\end{aligned} \tag{C.1}$$

The regressor of interest Last month$_{ifdys}$ is an indicator variable that equals one if a tender's i purchase was finalised at the last month of the fiscal year and zero otherwise. As in the benchmark specification 3 presented in section 3.1, we apply fixed effects at the level of the advertising department d, the bidding firm f, the year y and the industry of the purchased good s.

The key difference between the benchmark specification and the model above is that we separate the price ratio variable. We use the logarithm of the bid price as our dependent variable and control for the logarithm of the expected price. This specification is less restrictive than the benchmark specification of this paper.

We present our findings in Table C.5. Panel A presents results for all tenders while Panel B restricts the sample to winners only. All specifications control for the logarithm of the expected value and all fixed effects applied are listed at the bottom of the table. The findings suggest that firms participating in auctions bid between 1.9% and 1.2% higher prices in the last month of the fiscal year. The winning prices are between 0.9% and 0.6% higher in the last month, which is consistent with our previous estimates.

Table (C.5) The Year-End Effect for Winners and Losers with Alternative Specification

	Dependent variable:				
	Log Bid Value				
	(1)	(2)	(3)	(4)	(5)
Panel A: All lots					
Last Month	0.021***	0.016***	0.013***	0.012***	0.011***
	(0.001)	(0.0004)	(0.0004)	(0.0003)	(0.0003)
Log Expected Price	0.985***	0.987***	0.985***	0.986***	0.984***
	(0.0003)	(0.0002)	(0.0003)	(0.0003)	(0.0003)
Observations	14,477,78	14,477,783	14,477,783	14,477,783	14,477,783
R^2	0.992	0.993	0.994	0.994	0.994
Panel B: All winners					
Last Month	0.010***	0.008***	0.007***	0.007***	0.007***
	(0.0004)	(0.0004)	(0.0003)	(0.0003)	(0.0003)
Log Expected Price	0.991***	0.992***	0.989***	0.989***	0.990***
	(0.0003)	(0.0002)	(0.0004)	(0.0004)	(0.0003)
Observations	12,003,212	12,003,212	12,003,212	12,003,212	12,003,212
R^2	0.993	0.994	0.994	0.995	0.995
Year FE	Yes	Yes	Yes	Yes	Yes
Buyer FE	No	Yes	No	Yes	Yes
Firm FE	No	No	Yes	Yes	Yes
Industry FE	No	No	No	No	Yes

Note: The sample is composed of winners and losers of competitive and non-competitive lots. The dependent variable in all columns is the price ratio as defined in equation 4. Last month is a dummy variable that takes the value one if the tender was finalised in December and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are eight-digit sectoral fixed effects. The specification defined in equation 2 is presented in the last column. All coefficients should be interpreted as percentage changes to the price (relative to the expected price). Standard errors are in parentheses and are clustered by government departments. *p<0.1; **p<0.05; ***p<0.01.

C.6 Unconditional Quantile Regression

Finally, while the analysis in section 3.2 demonstrates how the conditional expectation of the price ratio changes at the end of the fiscal year, we use unconditional quantile regressions (Firpo, Fortin and Lemieux, 2009) to show how different quantiles are affected. We accomplish this in Table C.6 for the 10%, 25%, 50%, 75%, and 90% quantiles of the marginal distribution of the price ratio given the fixed effects and the last month dummy. We use the benchmark specification with all fixed effects applied with a sample of all competitive lots.[3]

Table (C.6) Unconditional Quantile Regression

	Price Ratio Quantile (Price Ratio Corresponding to this Quantile)				
	10% (67.8%)	25% (83.6%)	50% (96.3%)	75% (99.5%)	90% (100%)
	(1)	(2)	(3)	(4)	(5)
Last Month	1.140***	1.471***	1.014***	0.192***	0.127***
	(0.219)	(0.133)	(0.058)	(0.011)	(0.009)
Observations	870,898	870,898	870,898	870,898	870,898
R^2	0.308	0.373	0.423	0.388	0.418
Year FE	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes

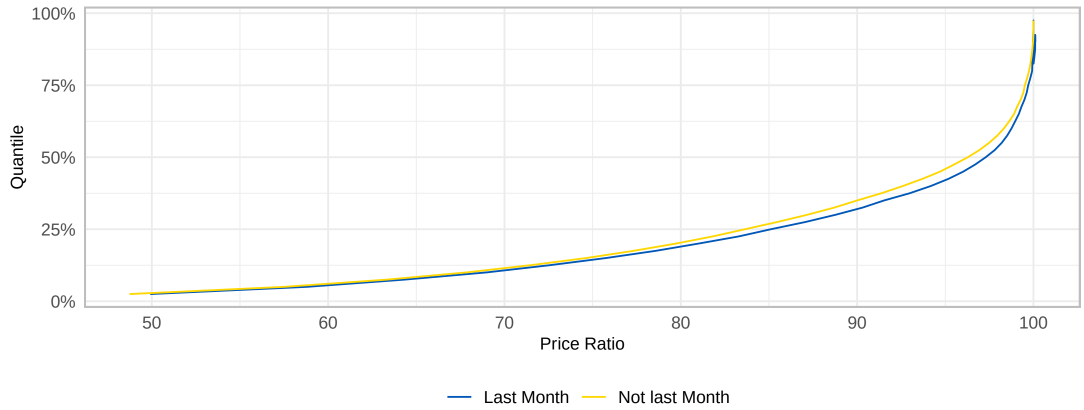
Note: The sample is composed of winners of competitive lots. The dependent variable in all columns is the price ratio as defined in equation 4. Last month is a dummy variable that takes the value one if the tender was finalised in December and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are eight-digit sectoral fixed effects. The specification defined in equation 2 is presented in the last column. All coefficients should be interpreted as percentage changes to the price (relative to the expected price). Standard errors are in parentheses and are clustered by government departments. *p<0.1; **p<0.05; ***p<0.01.

The coefficients in Table C.6 corresponding to each quantile are always positive but roughly decrease with higher quantiles. The values in this table can be interpreted as follows if there were a 1% increase in the proportion of lots completed in the last month, then the 10th unconditional quantile of the price ratio would increase by $1\% \times 1.140\% = 0.0114\%$.

Taking a finer grid of quantiles at 2.5% increments, we reproduce unconditional quantile regressions and plot the implied cumulative distribution of the last month lots against other month lots in Figure C.2. The findings demonstrate an increase in the proportion of lots that result in a price ratio of around one. That all coefficients are positive suggests that a marginal increase in the proportion of lots completed in the last month leads to higher prices across the spectrum with the prices of relatively cheap and relatively expensive purchases both increasing.

[3] We use a Gaussian kernel density in calculating the recentered influence function calculation. We restrict to competitive lots due to uncompetitive lots having a mass point on a price ratio of 1 which complicates the calculation of the recentered influence function.

Figure (C.2) Unconditional Quantile Regression Implied Cumulative Distributions



Note: The figure shows the unconditional quantile regression implied cumulative distributions of price ratios in the last month vs other months. It can be noted that a greater fraction of lots in the last month get a price ratio close to 100%.

D Institutional Context

Starting in 1997 Ukraine made a number of attempts to develop procurement legislation to harmonise it with the WTO and EU standards. However, these attempts have led to complicated legislation with a number of loopholes in a highly corrupted environment which discouraged the private sector from participating. According to the Ministry of Economic Development and Trade, 2.2% of Ukraine's GDP was lost annually due to inefficiency and corruption in public procurement (Transparency International Ukraine, 2017).

This situation persisted until the Euromaidan that took place from November 2013 to 2014. The protests, in response to the Ukrainian government suspending plans to more closely associate with the European Union, eventually resulted in the removal of President Yanukovich from power and spurred an unprecedented wave of reforms, including reforms to improve the procurement sector.

In 2014, a group of Euromaidan volunteers, the Ministry of Economic Development and Trade, and the private sector collaborated on the development of a new procurement platform with the goal to fight corruption in public procurement. The development was financially supported by the private sector, Transparency International Ukraine and international organisations. The procurement platform, named "Prozorro" (from the Ukrainian word for "transparent"), was launched in February 2015 with the motto, "Everyone sees everything." Prozorro aimed to change public procurement rules and make them more transparent by creating closer ties between the government, private sector, and civil society. The platform achieved a very high level of transparency by making government procurement transactions completely and freely available to the wider public. After the adoption of the new public procurement law on the 25th of December 2015 (Verkhovna Rada, 2015), the platform became mandatory for all public agencies from August 2016. Prozorro won major international government procurement awards such as World Procurement Award and Open Government Awards in 2016 and 2021 (Open Government Partnership, 2016, 2021; Procurement Leaders, 2016) and gained a showcase status for e-procurements awarded by the European Bank for Reconstruction and Development.

Alongside Prozorro, other mechanisms and supporting practices were initiated to improve other aspects of the procurement system. First, Transparency International Ukraine launched a public monitoring platform Dozorro in 2016. It allows anyone to report suspicious activities of government departments, participants or regarding the procure and submit complaints about any procurement. Second, Prozorro allows users to directly submit a complaint about discriminatory behaviour to the Anti-Monopoly Committee of Ukraine, the main competition regulator in Ukraine, which would investigate the case. Third, Prozorro launched a number of online courses describing the public procurement rules that are freely available to the public. Finally, Kyiv School of Economics launched the Centre of Excellence in Procurement. It offers training opportunities for both public as well as government officials and sellers and conducts research on public procurement.

While for the purposes of this paper Prozorro could be considered a unified marketplace, it is actually a confederation of 13 different commercial marketplaces. Most of these websites existed before Prozorro was launched and offer similar features with the main differences being appearances, APIs and the range of supporting services (such as legal services related to procurement) that they provide (Transparency International Ukraine, 2017). When a department wants to procure something they will post an advertisement on their favourite marketplace with duplicated advertisements being listed on the other 12 websites as well as Prozorro itself. Once this is done the job is visible to suppliers regardless of what commercial platform they are registered with. The rest of the procurement will be hosted by the commercial marketplace with the exception of auctions which take place on Prozorro itself. The commercial marketplace collects a fee for its services after a procurement job is awarded to a supplier. Finally, Prozorro has a business intelligence module which provides anyone with open access to all procurement details conducted on all platforms.

Regardless of the marketplace, under the implemented law there are several ways to purchase goods and services for a government entity. The first is the open procurement procedure which involves holding an auction which consists of several stages described in section 2.[4] The second

[4] Note that there are slight variations to the number of bidders and length of the procedure depending on the

approach is called competitive dialogue has two stages and can be applied in two situations: (1) if a government body can not determine the terms of the purchase (e.g. technical characteristics) and need to negotiate these before the tender or (2) if the subject of procurement is services (e.g. legal, software development), defining the requirements for which requires negotiations. In the first stage, the government authority discusses the terms and characteristics with potential sellers. Once a government agency is satisfied with proposals from sellers, the purchase moves to an auction stage which is identical to an auction held in an open procurement procedure.

The third negotiated procedure is the least competitive and allows the arrangement of a contract with the directly selected supplier. However, the government agency is required to submit a report about the purchase that will be available at Prozorro. During the negotiated procurement a government agency and a seller discuss terms and conditions before they reach an agreement. This procedure can only be applied to specific good categories (e.g. intellectual property, artwork) and in particular situations (e.g. absence of competition in a market, the tender procedure was cancelled twice due to an insufficient number of tenderers). Finally, for purchases that do not exceed 50,000 UAH in expected value, a government department can directly award a contract without holding a tender. In that case, a government department is required to release a report about such a purchase. This report should include details on the winning lot characteristics, the winner, expected value, winning values and the contract.

The procurement law does not apply to the purchases of the following classes of goods and services (Verkhovna Rada, 2015): production of banknotes, coins, state awards, official documents (e.g. passports, national identification); national security and defence; servicing and repayment of public debt; the management of gold and foreign exchange reserves; purchase, lease of land, buildings, other real estate or property; services of international arbitration courts and international commercial arbitrations for disputes involving a government agencies; services of financial organisations (e.g. provision of loans, guarantees, financial leasing and other financial services); financial services related to the issue, purchase, sale, transfer of securities or other financial instruments; services provided by the National Bank of Ukraine; scientific services awarded at competitive basis and defined by previous law; purchase of foreign diplomatic organisations; goods and services requested by the government ministries which ensures the formation and implementation of state policy in healthcare; electricity bought and sold on the electricity market; goods and services if their prices are approved by central government authorities.

expected value and choice of a government official conducting the purchase.

E The Model

Two models have previously been proposed in the literature, namely the precautionary savings model of Liebman and Mahoney (2017) and the procrastination model of Baumann (2019). The precautionary savings model considers government departments that face demand shocks each month. As a result, they create precautionary savings funds throughout the year. At the year-end, they cannot save these funds into the subsequent year and hence spend it all in the last month. The implication is that allowing departments to save funds between years would remove the incentive to spend quickly at the end of the fiscal year.

Two key reasons suggest, however, that other factors may be at work aside from precautionary savings. The first is that the UK has allowed government departments to save across fiscal years which has not noticeably decreased end of fiscal year spending. The second reason is that it is not clear if the month-to-month uncertainty required to explain the size of spending seen in practice is credible. As a result of these concerns Baumann (2019) suggested procrastination as a driving factor. Government departments typically have fiscal and performance reporting on a fiscal year basis. Budgetary managers in government departments need to incur effort costs in order to spend funds. As a result of these factors, they may backload effort until later in the fiscal year which results in the end of the fiscal year spending spikes.

Ukrainian government departments (and most world government departments) face stochastic costs of procurement. This suggests a third mechanism that drives end of fiscal year spending in this context. Our *stochastic costs* model is driven by government departments which in the early months of the year choose projects with considerable uncertainty as to the final cost. Later in the year departments would prefer to avoid these risky projects as they are faced with the possibility of not being able to afford expensive procurements or having funds left over if projects come in cheaper than expected. As a result, they have an incentive to switch to less risky projects that have a higher expected cost. The behaviour we can see in the Ukrainian context, where departments choose more and smaller procurement jobs at the end of the fiscal year may be a result of these incentives. While more and smaller jobs are likely to face higher contracting costs the procurement risk is more diversified.

There is some evidence of this occurring in the UK, where Hyndman, Jones and Pendlebury (2007) found that a popular tactic amongst managers was to maintain a list of off the shelf projects that could be used if there was no better use of the funds before the end of the fiscal year. We can clearly see in the Ukrainian data that departments are much more likely to engage in uncompetitive bargaining procedures in the final months of the fiscal year. In addition, they are more likely to split jobs so that there are several small jobs that can be negotiated directly with suppliers rather than one large job that requires a competitive tender process.

The main agent in our model is a government department. A department chooses an expected value, x, for procurement in each period. The realised price, p, of the procurement, can differ from the expected value. There are two types of projects that can be undertaken. The first has a random cost where with 50% probability, the price is equal to x while with 50% probability the price is θx where $\theta < 1$. The second has a deterministic cost where the price will be $x\gamma$ for $\frac{1+\theta}{2} < \gamma < 1$ with certainty. Therefore, the deterministic project has a higher expected price than the random cost project but is cheaper than if the random cost project is expensive. The department can divide its spending between the two types of projects each month.

A department cannot engage in a procurement where it is possible that they will not be able to pay in the event of a high price.[5] In each month they get a utility equal to an increasing concave function of the expected value, $u(x_m) = x_m^\delta$. The department has an annual budget normalised to 1 that they can spend over a year.

This setting can be summarized as:

$$V_m(B_m) = \max_{d_m \in [0,1], x_m \in [0, \frac{B_m}{d_m \gamma + (1-d_m)}]} [u(x_m) + \beta V_{m'}(B_{m'}(\bar{x}_m))] \tag{E.1}$$

$$\text{Where} \qquad \bar{x}_m = d_m \gamma x_m + (1 - d_m) \times (x_m \quad \text{or} \quad \theta x_m)$$

$$B_{m'}(\bar{x}_m) = 1 + \lambda(B_{m'} - \bar{x}_m) \qquad \text{if } m \equiv M$$

$$= B_{m'} - \bar{x}_m \qquad \text{if } m \not\equiv M$$

Where the annual budget is 1, M is the number of months in the year, λ is the amount of rollover that is allowed, m and m' are the indices for this month and next month respectively. d_m is the proportion of monthly spending devoted to deterministic projects while x_m is the total expected spending.

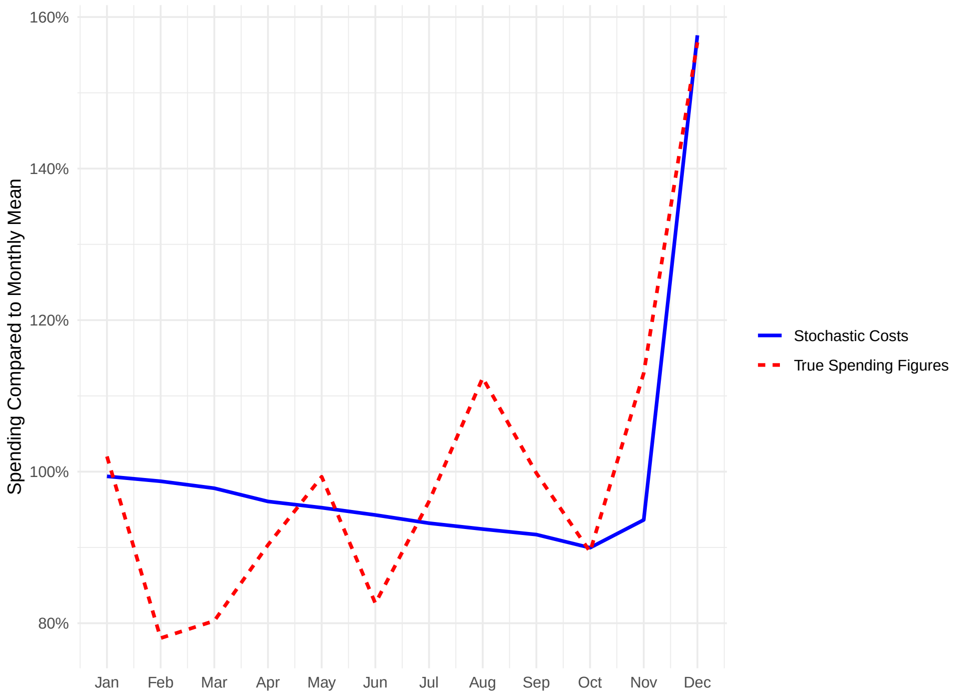
This model is a standard consumption smoothing problem, and with a suitable parameterisation, it exhibits end of fiscal year spending spikes. As the prices paid in early months are stochastic and leave an uncertain amount for later months, the department will spend less in early months leading to precautionary savings. Later in the year, there is more chance of having leftover funds that may not be spent if random projects are selected. Hence the department switches to deterministic projects. As there is no more uncertainty in procurement prices and

[5] In a real setting any price is possible and it is, therefore, possible that a department will need to pull out of a procurement for cost reasons. This simple model captures the essential behaviour, however, that departments face stochastic costs in procurements and need to budget for these stochastic costs.

no way to save precautionary savings between fiscal years, these savings are spent, which results in the heightened end of fiscal year spending.

We calibrate this model to the Ukrainian spending series in the following way. First, we calibrate θ such that the distribution of price ratios in the model matches those in our data. This results in a figure of $\theta = 0.545$. Second, we then impose that deterministic projects are 2% more expensive (in expectation) than the stochastic project.[6] This corresponds to a value of $\gamma = 0.796$. We impose $\beta = 0.996$ which implies 5% annual discounting. Third, we calibrate the curvature of the utility function so that the ratio of last month to year average spending equals the ratio we observe in the data (see table 1). This implies a value of $\delta = 0.58$. We present the spending profile that this generates over the year in figure E.1. In this case, in the early months of the fiscal year, the department purely engages in the risky project. In the later months of the year, the department shifts to a greater fraction of spending on the deterministic project. In the final month of the year, the department purely spends in a deterministic fashion.

Figure (E.1) Fit of Model Implied Month Spending Averages to Ukrainian Procurement Data Monthly Averages



Note: The grey shape in the background represents the monthly average spending amounts for the total Ukrainian public sector over the years of our study. The line in the foreground shows the monthly spending averages implied by our model once calibrated.

This mechanism has some similarities with the precautionary savings model of Liebman and

[6]This is in line with the regression results presented earlier in this paper. This low value also highlights that the model does not depend on an exogenous increase (from higher θ) in costs to drive end of fiscal year spending. End of fiscal year spending is not purely driven by engaging in projects that are more expensive in expectation, but rather by how departments change their behaviour between risky and deterministic projects.

Mahoney (2017). Both models are driven by uncertainty regarding future marginal utility from spending. Similarly to that model, allowing departments to roll over unspent funds prevents the end of fiscal year spending spikes. There are several key differences, however, when we consider the stochastic costs model. The first difference is that the uncertainty comes from procurement prices rather than demand shocks, which is a more measurable and, for some government departments and procurement contexts, more prevalent source of uncertainty. In the calibration of the model, it was possible to base the parameters governing the variability of cost on data observables. This is particularly important as it was unclear if the volatility required for the Liebman-Mahoney calibration to fit the data was credible given the uncertainty in demand faced by real government departments.

The second is that the stochastic costs mechanism requires the availability of a risk-free way to spend funds. In the absence of such an option end of the fiscal year, spending is prevented due to the necessity to hold funds in reserve in case a project comes in at an expensive price. While departments build up a precautionary savings fund and would like to spend it at the end of the year, this requirement prevents them. The third is that new policy implications become more apparent.

One new policy implication is that spending spikes may be assuaged by reducing the procurement price risk that departments face. This may be achievable through a centralised insurance system where departments that got many bidders on their projects make transfer payments to departments that had fewer bidders participating in their auctions. With the current calibration, if the uncertainty in the risky project is reduced, so that $\theta = 0.6$, with a corresponding increase in γ and so that it remains 2% above the expectation of the risky project, we get the final month spending at 1.5 times the monthly average. Hence while there is still the end of fiscal year spending, there is less than in the calibrated case (where last month's spending is 1.57 times higher). While insurance is in principle a solution to the original Liebman-Mahoney model, there is the practical problem that demand shocks are less observable and thus harder for departments to agree on insurance arrangements against them.

The second policy implication is that end of fiscal year spending may be averted by cutting back on the department's opportunities for deterministic procurements at the end of the fiscal year. This approach is to some extent counterintuitive as having a list of ready to go projects have been suggested as an effective way to respond to fiscal year budgeting constraints (Hyndman, Jones and Pendlebury, 2007). This may involve restrictions on departments engaging in reported procurements at the year-end in a Ukrainian context. Alternatively, a government might be able to deter departments by effectively charging them a tax (or in alternative terms a budgetary reduction) from engaging in deterministic projects. In the Ukrainian context, deterministic projects may be considered to be all procurements that do not involve a competitive auction process. As an example, if they engaged in an uncompetitive tender project at the end of the fiscal year and spent an amount to money x then ψx (for some $\psi > 1$) would be deducted from their budget with $(\psi - 1)x$ of this money being directly returned to treasury.[7]

To observe this, we take the benchmark calibration and increase γ by 10% so that it is 0.867. The result is that end of fiscal year spending is deterred significantly relative to the benchmark case with final month spending being 1.31 times more than the average month. While this example shows that governments charging taxes to departments may deter end of fiscal year spending, it also demonstrates that heightened prices from suppliers at the end of the fiscal year may themselves deter end of fiscal year spending.

Finally and similarly to the Liebman and Mahoney (2017) model, allowing departments to roll over funds between fiscal years would prevent the end of fiscal year spending. While uncertainty in costs would still drive departments to build up precautionary savings, there would not be any shift to spend these funds at the end of the year as this precautionary savings fund could be carried over to the following year. This policy implication has some practical limitations, however. The first is that rollover was implemented in the United Kingdom and did not lead to any discernible decrease at the end of fiscal year spending (Baumann, 2019). This may imply that other forces are leading to the end of fiscal year spending in the UK and potentially also in Ukraine. For instance, stochastic costs may not be the only mechanism leading to end of fiscal year spending. The second complication is that the central government allowing rollover will not be felt by lower level budgetary managers if their "savings" are taken from them by higher level budgetary managers. It is not clear if this would be a problem in the Ukrainian public sector.

E.1 Two period version of model

We consider a two period version of the model introduced in section E. Rather than a concrete utility function, we instead consider an arbitrary increasing concave utility function $f(x)$ and consider the case where there is no rollover ($\lambda = 0$).

We first establish that the department will spend its entire budget using deterministic projects in the second and final period of the year.

[7]This is similar to the time variant budgetary tax policy recommendation of Baumann (2019) with the difference being these taxes are levied based on the competitiveness of the purchase rather than when in the year it was undertaken. This may also be simpler to understand and implement than time variant taxes.

Lemma 1. *The department will spend its entire remaining budget using deterministic projects in the last period of the fiscal year.*

Proof. Utility is increasing in the total expected value, x_E, that is spent. As the department needs to have enough funds to be able to cover expenses if costs end up being expensive the constraint for the department is $d\gamma x_E + (1-d)x_E < B_2$ where B_2 is the budget available in the second period of the year. As $\gamma < 1$ the total spend is maximized when $d = 1$ and all funds are spent with deterministic projects.

Now we can derive an expression for the maximisation problem for a department at the start of the first period:

$$\max_{d\in[0,1],x\in[0,\frac{B_1}{d\gamma+(1-d)}]} f(x) + \frac{\beta}{2}\left[\overbrace{f(B_1 - d\gamma x - (1-d)\theta x)}^{\text{If Stochastic Project Cheap}} + \overbrace{f(B_1 - d\gamma x - (1-d)x)}^{\text{If Stochastic Project Expensive}}\right] \tag{E.2}$$

Given this, we can show that there exist parameterisations, such that the department will, in expectation, spend more in the second round than in the first round.

Using the parameters $\beta = 0.975$, $\theta = 0.5$, $\gamma = 0.8$, $B = 1$ and a log utility function $f(x) = \log(x)$ we can numerically solve for the department spending a fraction $d = 0.34$ in the first period. The expected value in the first period is 0.633. This implies that in expectation 0.486 will be spent in the first period, and the remaining 0.514 will be spent deterministically in the second period.

F Empirical Robustness Checks

This section tests the robustness of our empirical findings. First, in Table F.1, we present our standard specifications using a dummy for each month of the calendar year instead of only including the last month dummy. The first column presents the results without conditioning on the number of bidders, while the second column includes this conditioning. The results show that the price ratio is highest at year-end.

Second, the levels of inflation in Ukraine varied a lot throughout our study period. While we apply the monthly price index throughout the entire analysis, we also replicate our specification for each fiscal year separately in Table F.2. Our findings demonstrate that the last month effect is present in years with both high and low inflation levels.

Third, we demonstrate that our benchmark results are robust to different standard error structures in Table F.3. Specifically, we use robust standard errors, clustering at buyer, seller, industry, buyer and seller levels, as well as bootstrapping.

Next, in Table F.4, we show that our findings are robust to alternative fixed effects structures, such as combined year×buyer and year×firm fixed effects. In Table F.5, we replace the last month dummy with the last two weeks of year-end dummy.

Fifth, we plot daily busy-ness and outside option values in Figure F.1. Table F.6 presents our main results without controlling for outside option and busy-ness, where the last month dummy increases slightly compared to the benchmark table.

Sixth, in Table F.7, we examine whether competition changes at the fiscal year-end. To test this, we regress the number of participations (in logarithm) on the last month. The results indicate that competition is about 2%-3% lower at fiscal year-end.

Seventh, in Table F.8, we replicate our benchmark specification while controlling for whether the procedure is competitive. The results show that competitive procedures are, on average, about 8% cheaper than non-competitive procedures.

Next, in the main body of the paper, we used a probit model to estimate the likelihood of a risky firm participating and winning a tender at year-end. We replicate all these specifications using a linear probability model and present our findings in Table F.9.

Finally, Table F.10 performs a similar exercise but presents costs and margins as estimated in our benchmark specification. For scaling, the costs are divided by the expected price, while margins are calculated by dividing bids by inferred costs.[8] We do not use margin common cost

[8]Expected price is not reflected in margin calculations. Given that prices are divided by costs, scaling issues

and exclude the costs and margins of the last bidders in each auction and an observation here corresponds to a seller-lot. These factors mean that these results cannot be directly compared to earlier results. The results do not show clear pattern in costs. Costs are lowest at the start of the year, highest in the middle, and moderate at year-end. One possible interpretation for seeing lower costs at year-end is that the expected price could increase in anticipation of higher prices, which would affect the ratio of cost to expected price. However, it is reassuring that our margin increases are not driven by cost increases at year-end but by the prices bid by firms. Costs are used as an alternative benchmark of how much a job "should" cost, but it is not the changes in this benchmark that is driving our results.

are avoided.

Table (F.1) The Year-End Effect by Month with Bidders

	Dependent variable: Price Ratio	
	(1)	(2)
Last Month	1.086***	0.215***
	(0.028)	(0.017)
February	0.162***	0.015
	(0.023)	(0.016)
March	0.277***	0.012
	(0.026)	(0.018)
April	0.432***	0.046**
	(0.028)	(0.020)
May	0.414***	0.079***
	(0.026)	(0.019)
June	0.527***	0.100***
	(0.026)	(0.018)
July	0.568***	0.143***
	(0.026)	(0.018)
August	0.584***	0.129***
	(0.027)	(0.018)
September	0.680***	0.180***
	(0.026)	(0.018)
October	0.736***	0.213***
	(0.026)	(0.018)
November	0.729***	0.161***
	(0.027)	(0.018)
Observations	11,967,824	11,967,824
R^2	0.332	0.538
Number Bidders	No	Yes
Outside options/Busyness	No	Yes
Year FE	Yes	Yes
Buyer FE	Yes	Yes
Firm FE	Yes	Yes
Industry FE	Yes	Yes

Note: The sample is composed of winners of competitive and non-competitive lots. The dependent variable in all columns is the price ratio as defined in equation 4. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are eight-digit sectoral fixed effects. All coefficients should be interpreted as percentage changes to the price (relative to the expected price). Standard errors are in parentheses and are clustered by government departments. *p<0.1; **p<0.05; ***p<0.01.

Table (F.2) The Year-End Effect by Fiscal Year and Months

	Dependent variable: Price Ratio				
	(1)	(2)	(3)	(4)	(5)
	FY 2017	FY 2018	FY 2019	FY 2020	FY 2021
Last Month	4.251***	2.808***	1.863***	2.044***	1.817***
	(0.127)	(0.095)	(0.079)	(0.076)	(0.045)
February	0.231*	−0.183**	0.030	0.217***	0.690***
	(0.121)	(0.090)	(0.078)	(0.081)	(0.045)
March	0.563***	0.088	0.241***	1.235***	0.747***
	(0.130)	(0.094)	(0.083)	(0.081)	(0.049)
April	1.220***	0.583***	0.445***	1.277***	0.926***
	(0.137)	(0.102)	(0.084)	(0.085)	(0.050)
May	1.582***	0.786***	0.462***	0.494***	0.989***
	(0.135)	(0.092)	(0.085)	(0.081)	(0.050)
June	1.973***	1.086***	0.704***	0.876***	1.132***
	(0.136)	(0.098)	(0.083)	(0.079)	(0.049)
July	2.353***	1.405***	0.600***	0.991***	1.209***
	(0.139)	(0.098)	(0.082)	(0.077)	(0.049)
August	2.478***	1.524***	0.463***	1.119***	1.107***
	(0.137)	(0.097)	(0.092)	(0.078)	(0.050)
September	2.724***	1.711***	0.461***	1.310***	1.286***
	(0.130)	(0.097)	(0.081)	(0.077)	(0.048)
October	2.914***	1.745***	0.654***	1.426***	1.318***
	(0.130)	(0.096)	(0.084)	(0.077)	(0.048)
November	3.439***	1.684***	0.460***	1.356***	1.378***
	(0.131)	(0.097)	(0.087)	(0.078)	(0.046)
Observations	1,239,745	1,465,369	1,601,040	3,976,512	5,539,998
R^2	0.495	0.484	0.500	0.454	0.408
Buyer FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes

Note: The sample is composed of winners of competitive and non-competitive lots. The dependent variable in all columns is the price ratio as defined in equation 4. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are eight-digit sectoral fixed effects. All coefficients should be interpreted as percentage changes to the price (relative to the expected price). Standard errors are in parentheses and are clustered by government departments. *p<0.1; **p<0.05; ***p<0.01.

Table (F.3) Benchmark Results with Different Standard Errors

	Dependent variable:				
	Price Ratio				
	(1)	(2)	(3)	(4)	(5)
Panel A: All lots					
Last Month	0.718***	0.542***	0.462***	0.428***	0.423***
[Robust SE]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
[Buyer SE]	[0.031]	[0.025]	[0.019]	[0.017]	[0.017]
[Seller SE]	[0.025]	[0.023]	[0.017]	[0.015]	[0.015]
[Buyer & Seller SE]	[0.031]	[0.025]	[0.019]	[0.017]	[0.017]
[Industry SE]	[0.078]	[0.070]	[0.049]	[0.045]	[0.044]
[Bootstrapped SE]	[0.004]	[0.004]	[0.003]	[0.003]	[0.003]
Observations	11,967,900	11,967,900	11,967,900	11,967,900	11,967,900
R^2	0.023	0.141	0.290	0.329	0.332
Panel B: All lots conditioning on number of bidders					
Last Month	0.115***	0.117***	0.078***	0.084***	0.074***
[Robust SE]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
[Buyer SE]	[0.024]	[0.021]	[0.015]	[0.015]	[0.015]
[Seller SE]	[0.020]	[0.019]	[0.013]	[0.013]	[0.013]
[Buyer & Seller SE]	[0.023]	[0.021]	[0.015]	[0.015]	[0.015]
[Industry SE]	[0.052]	[0.052]	[0.039]	[0.039]	[0.037]
[Bootstrapped SE]	[0.005]	[0.004]	[0.004]	[0.004]	[0.004]
Observations	11,967,900	11,967,900	11,967,900	11,967,900	11,967,900
R^2	0.428	0.461	0.521	0.537	0.539
Year FE	Yes	Yes	Yes	Yes	Yes
Buyer FE	No	Yes	No	Yes	Yes
Firm FE	No	No	Yes	Yes	Yes
Industry FE	No	No	No	No	Yes

Note: The sample is composed of winners of competitive and non-competitive lots. The dependent variable in all columns is the price ratio as defined in equation 4. Last month is a dummy variable that takes the value one if the tender was finalised in December and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are eight-digit sectoral fixed effects. The specification defined in equation 2 is presented in the last column. All coefficients should be interpreted as percentage changes to the price (relative to the expected price). Standard errors are in square brackets. *p<0.1; **p<0.05; ***p<0.01.

Table (F.4) The Year-End Effect with Alternative Fixed Effects

	Dependent variable:				
	Price Ratio				
	(1)	(2)	(3)	(4)	(5)
Panel A: All lots					
Last Month	0.718***	0.498***	0.425***	0.385***	0.378***
	(0.019)	(0.013)	(0.010)	(0.009)	(0.009)
Observations	11,967,900	11,967,900	11,967,900	11,967,900	11,967,900
R^2	0.024	0.195	0.372	0.425	0.427
Panel B: All lots conditioning on number of bidders					
Last Month	0.115***	0.114***	0.077***	0.084***	0.073***
	(0.016)	(0.012)	(0.009)	(0.009)	(0.008)
Busyness	0.039***	0.005	0.028***	0.012***	0.018***
	(0.010)	(0.006)	(0.004)	(0.003)	(0.003)
Outside Option	−0.017***	−0.013***	−0.001	−0.001	0.002
	(0.006)	(0.004)	(0.003)	(0.003)	(0.003)
2 bidders	−9.304***	−8.682***	−8.996***	−8.727***	−8.859***
	(0.113)	(0.075)	(0.061)	(0.056)	(0.056)
3 bidders	−17.757***	−16.755***	−16.066***	−15.688***	−15.769***
	(0.129)	(0.087)	(0.080)	(0.069)	(0.069)
4 bidders	−22.958***	−21.838***	−20.620***	−20.191***	−20.264***
	(0.156)	(0.109)	(0.103)	(0.089)	(0.088)
5 bidders	−27.360***	−26.094***	−24.472***	−23.980***	−24.046***
	(0.181)	(0.137)	(0.124)	(0.112)	(0.110)
6 + bidders	−32.737***	−31.452***	−29.256***	−28.737***	−28.807***
	(0.256)	(0.179)	(0.152)	(0.134)	(0.130)
Observations	11,967,900	11,967,900	11,967,900	11,967,900	11,967,900
R^2	0.427	0.482	0.564	0.591	0.593
Year FE	Yes	No	No	No	No
Buyer-Year FE	No	Yes	No	Yes	Yes
Firm-Year FE	No	No	Yes	Yes	Yes
Industry FE	No	No	No	No	Yes

Note: The sample is composed of winners of competitive and non competitive lots. The dependent variable in all columns is the price ratio as defined in equation 4.Last month is a dummy variable that takes the value one if the tender was finalised in December and zero otherwise. Year FE stands for year fixed effects, buyer-year FE are government departments and year combined fixed effects, firm-year FE are firm and year combined fixed effects, and industry FE are eight-digit sectoral fixed effects. All coefficients should be interpreted as percentage changes to the price (relative to the expected price). Standard errors are in parentheses and are clustered by government departments. *p<0.1; **p<0.05; ***p<0.01.

Table (F.5) The Two Weeks Year-End Effect

	Dependent variable:				
	Price Ratio				
	(1)	(2)	(3)	(4)	(5)
Panel A: All lots					
Last Two Weeks	0.824^{***}	0.604^{***}	0.485^{***}	0.449^{***}	0.443^{***}
	(0.022)	(0.017)	(0.013)	(0.013)	(0.013)
Observations	11,967,900	11,967,900	11,967,900	11,967,900	11,967,900
R^2	0.024	0.141	0.290	0.329	0.332
Panel B: All lots conditioning on number of bidders					
Last Two Weeks	0.048	0.060^{*}	0.024	0.039^{*}	0.036^{*}
	(0.034)	(0.031)	(0.021)	(0.021)	(0.020)
Busyness	0.054^{***}	0.018^{**}	0.042^{***}	0.026^{***}	0.031^{***}
	(0.011)	(0.008)	(0.006)	(0.005)	(0.005)
Outside Option	-0.024^{**}	-0.020^{**}	-0.006	-0.005	0.002
	(0.010)	(0.008)	(0.006)	(0.006)	(0.006)
2 bidders	-9.307^{***}	-8.947^{***}	-9.124^{***}	-8.940^{***}	-9.082^{***}
	(0.163)	(0.133)	(0.128)	(0.116)	(0.116)
3 bidders	-17.761^{***}	-17.120^{***}	-16.453^{***}	-16.167^{***}	-16.251^{***}
	(0.185)	(0.156)	(0.155)	(0.141)	(0.140)
4 bidders	-22.962^{***}	-22.230^{***}	-21.149^{***}	-20.824^{***}	-20.898^{***}
	(0.217)	(0.184)	(0.178)	(0.164)	(0.163)
5 bidders	-27.363^{***}	-26.514^{***}	-25.114^{***}	-24.729^{***}	-24.792^{***}
	(0.243)	(0.210)	(0.195)	(0.183)	(0.180)
6 or more bidders	-32.741^{***}	-31.864^{***}	-30.051^{***}	-29.652^{***}	-29.719^{***}
	(0.327)	(0.271)	(0.228)	(0.214)	(0.210)
Observations	11,967,900	11,967,900	11,967,900	11,967,900	11,967,900
R^2	0.427	0.460	0.520	0.536	0.539
Year FE	Yes	Yes	Yes	Yes	Yes
Buyer FE	No	Yes	No	Yes	Yes
Firm FE	No	No	Yes	Yes	Yes
Industry FE	No	No	No	No	Yes

Note: The sample is composed of winners of competitive and non-competitive lots. The dependent variable in all columns is the price ratio as defined in equation 4. Last two weeks is a dummy variable that takes the value one if the tender was finalised after December 15th and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are eight-digit sectoral fixed effects. All coefficients should be interpreted as percentage changes to the price (relative to the expected price). Standard errors are in parentheses and are clustered by government departments. *p<0.1; **p<0.05; ***p<0.01.

Table (F.6) The Year-End Effect Excluding Busyness and Outside Option

	Dependent variable:				
	Price Ratio				
	(1)	(2)	(3)	(4)	(5)
Last Month	0.158***	0.126***	0.108***	0.099***	0.093***
	(0.009)	(0.008)	(0.007)	(0.007)	(0.007)
2 bidders	−9.305***	−8.944***	−9.121***	−8.937***	−9.079***
	(0.113)	(0.077)	(0.063)	(0.059)	(0.059)
3 bidders	−17.759***	−17.117***	−16.450***	−16.164***	−16.249***
	(0.129)	(0.090)	(0.084)	(0.073)	(0.073)
4 bidders	−22.960***	−22.227***	−21.147***	−20.821***	−20.897***
	(0.156)	(0.110)	(0.107)	(0.094)	(0.093)
5 bidders	−27.363***	−26.510***	−25.113***	−24.727***	−24.792***
	(0.181)	(0.138)	(0.128)	(0.117)	(0.115)
6 + bidders	−32.744***	−31.860***	−30.054***	−29.652***	−29.722***
	(0.257)	(0.184)	(0.161)	(0.144)	(0.139)
Observations	11,967,900	11,967,900	11,967,900	11,967,900	11,967,900
R^2	0.427	0.460	0.520	0.536	0.539
Year FE	Yes	Yes	Yes	Yes	Yes
Buyer FE	No	Yes	No	Yes	Yes
Firm FE	No	No	Yes	Yes	Yes
Industry FE	No	No	No	No	Yes

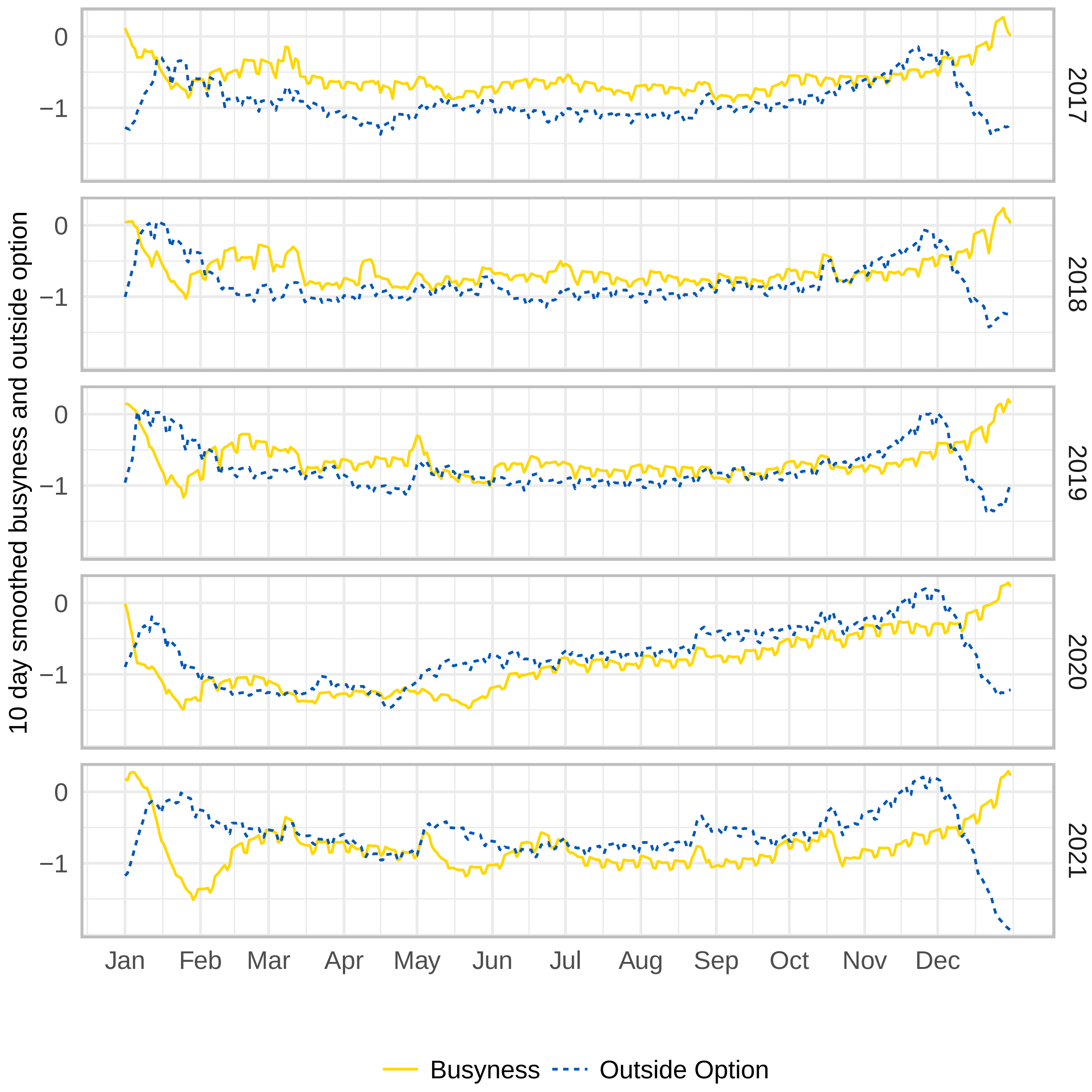
Note: The sample is composed of winners of competitive and non-competitive lots. The dependent variable in all columns is the price ratio as defined in equation 4. Last month is a dummy variable that takes the value one if the tender was finalised in December and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are eight-digit sectoral fixed effects. All coefficients should be interpreted as percentage changes to the price (relative to the expected price). Standard errors are in parentheses and are clustered by government departments. *p<0.1; **p<0.05; ***p<0.01.

Table (F.7) Competition at the Fiscal Year-End

	Dependent variable:				
	Ln Number of Participants				
	(1)	(2)	(3)	(4)	(5)
Last Month	-0.036^{***}	-0.028^{***}	-0.023^{***}	-0.021^{***}	-0.021^{***}
	(0.001)	(0.0005)	(0.0004)	(0.0003)	(0.0003)
Observations	12,003,212	12,003,212	12,003,212	12,003,212	12,003,212
R^2	0.038	0.189	0.367	0.410	0.418
Year FE	Yes	Yes	Yes	Yes	Yes
Buyer FE	No	Yes	No	Yes	Yes
Firm FE	No	No	Yes	Yes	Yes
Industry FE	No	No	No	No	Yes

Note: The sample is composed of winners of competitive and non-competitive lots. The dependent variable in all columns is the logarithm of tender participants. Last month is a dummy variable that takes the value one if the tender was finalised in December and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are eight-digit sectoral fixed effects. The specification defined in equation 2 is presented in the last column. All coefficients should be interpreted as percentage changes to the price (relative to the expected price). Standard errors are in parentheses and are clustered by government departments. *p<0.1; **p<0.05; ***p<0.01.

Figure (F.1) Busy-ness and Outside Option Over the Year



Note: This figure shows indices for the busy-ness (how many lots have been awarded in the preceding 30 days) and the outside option (how many lots will be awarded in the next 30 days) for each month and year. The outside option is high in the lead-up to the end of the fiscal year before it levels off as a result of Christmas and the holidays. Suppliers remain busy completing lots that were finalised at this time.

Table (F.8) The Year-End Effect Controlling for Competitive Procedures

	(1)	(2)	(3)	(4)
Panel A: All lots				
Last mont	1.101***	0.770***	0.684***	0.587***
	(0.033)	(0.022)	(0.017)	(0.015)
Competitive	−9.657***	−9.154***	−7.552***	−7.962***
	(0.114)	(0.093)	(0.083)	(0.080)
Observations	14,424,703	14,424,703	14,424,703	14,424,703
R^2	0.162	0.262	0.396	0.431
Panel B: All lots conditioning on number of bidders				
Last month	0.153***	0.149***	0.119***	0.062***
	(0.024)	(0.018)	(0.014)	(0.013)
Competitive	0.907***	−0.096	−0.622***	−1.173***
	(0.112)	(0.072)	(0.066)	(0.056)
Observations	14,424,477	14,424,477	14,424,477	14,424,477
R^2	0.425	0.451	0.530	0.546
Year FE	Yes	Yes	Yes	Yes
Buyer FE	No	Yes	No	Yes
Seller FE	No	No	Yes	Yes
Industry FE	No	No	No	Yes

Note: The sample is composed of winners and losers of competitive and non-competitive lots. The dependent variable in all columns is the price ratio as defined in equation 4. Last month is a dummy variable that takes the value one if the tender was finalised in December and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are eight-digit sectoral fixed effects. The specification defined in equation 2 is presented in the last column. All coefficients should be interpreted as percentage changes to the price (relative to the expected price). Standard errors are in parentheses and are clustered by government departments. *p<0.1; **p<0.05; ***p<0.01.

Table (F.9) Year End Effect for Risky Suppliers

	Dependent variable (DV):		
	(1)	(2)	(3)
Panel A: DV Risky participant			
Last month	0.036^{***}	0.024^{***}	0.013^{***}
	(0.007)	(0.005)	(0.003)
Observations	2,039,715	2,039,715	2,039,715
R^2	0.008	0.076	0.137
Panel B: DV Risky winner			
Last month	0.027^{***}	0.016^{***}	0.008^{***}
	(0.005)	(0.004)	(0.003)
Observations	874,357	874,357	874,357
R^2	0.008	0.129	0.195
Year FE	Yes	Yes	Yes
Buyer FE	No	Yes	Yes
Industry FE	No	No	Yes

Note: The sample in panel A is composed of winners and losers of competitive lots, while in panel B the sample is restricted to winners of competitive lots. The dependent variable in all columns in panel A is whether a risky firm is a participant and in panel B is whether a risky firm is a winner. Last month is a dummy variable that takes the value 1 if the tender was finalised in December and 0 otherwise. Year FE stands for year fixed effects, buyer FE are government departments' fixed effects, and industry FE are eight-digit sectoral fixed effects. Standard errors are in parentheses and are clustered by supplier level *p<0.1; **p<0.05; ***p<0.01.

Table (F.10) The Year End Effect for Margins and Costs

	Dependent variable:			
	Implied Cost		Margint	
	(1)	(2)	(3)	(4)
Last Month	0.218*	0.504***	1.353***	0.588*
	(0.130)	(0.136)	(0.310)	(0.344)
February	0.475***	−0.177	−0.745***	0.404
	(0.117)	(0.119)	(0.279)	(0.301)
March	1.188***	0.096	−1.950***	0.016
	(0.117)	(0.122)	(0.280)	(0.309)
April	1.751***	0.157	−2.751***	0.176
	(0.117)	(0.124)	(0.280)	(0.312)
May	1.882***	0.585***	−3.338***	−0.742**
	(0.118)	(0.125)	(0.283)	(0.316)
June	1.967***	0.515***	−3.044***	−0.272
	(0.118)	(0.125)	(0.281)	(0.315)
July	2.451***	0.897***	−3.606***	−0.641**
	(0.115)	(0.123)	(0.275)	(0.309)
August	2.402***	1.021***	−3.411***	−0.924***
	(0.117)	(0.125)	(0.280)	(0.315)
September	2.521***	1.386***	−3.579***	−1.564***
	(0.115)	(0.122)	(0.274)	(0.309)
October	2.645***	1.445***	−3.913***	−1.707***
	(0.114)	(0.122)	(0.273)	(0.308)
November	1.791***	0.602***	−2.171***	0.301
	(0.111)	(0.119)	(0.266)	(0.301)
Observations	368,477	368,461	368,477	368,461
R^2	0.009	0.435	0.004	0.367
Year FE	Yes	Yes	Yes	Yes
Buyer FE	No	Yes	No	Yes
Firm FE	No	Yes	No	Yes
Sector FE	No	Yes	No	Yes

Note: The sample is composed of winners and losers of competitive lots for which we have estimated a cost. These include winning and losing bids, excluding the last bidder in the final round. We have not taken the median over costs for each job so these results cannot be directly compared to other costs where this is done to get a baseline value for the job. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are two digit sectoral fixed effects. All coefficients should be interpreted as percentage changes to the price (relative to the expected price). Standard errors are in parentheses and are clustered by government departments. *p<0.1; **p<0.05; ***p<0.01.

G Neural Network Monte Carlo

We performed a Monte Carlo exercise to validate our neural networks approach for backing costs out of bids. The most clearcut way to do this would be to simulate the full auction procedure. We are not able to do this as we are not able to solve for optimal strategies of bidders in the auction.[9] As a result, we have taken a simpler approach that allows us to test the neural network approach.

We generate a "Monte Carlo dataset" with 814408 observations and 8 variables - the same number as we have in our "real dataset". The 8 variables, denoted as $\{v_k\}_{k=1:8}$, are normally distributed and drawn from a multivariate Gaussian distribution with a mean of zero and a specified covariance matrix. We have three versions of the Monte Carlo, one of which has the covariance matrix being the identity matrix, one of which has the covariance matrix being drawn randomly from the inverse Wishart distribution and a third when the covariance matrix is an equal mix of the identity matrix and a random matrix from the inverse Wishart distribution. These 8 variables are going to undertake the same preprocessing as variables in our main specification including a z score transformation so that each is $N(0,1)$ and then used in place of the neural network inputs in figure 5.

We then create a deterministic mapping from each set of these 8 variables to an (n_1, n_2, n_3) tuple from a particular observation in the real dataset tuple. This ensures that our Monte Carlo dataset will have the same distribution of $(n_1, n_2, n_3, p, \tilde{p})$ as our real dataset. To achieve this:

1. We sort the real dataset in ascending order according to the estimated demand as calculated via equation 6. This is done as an ad-hoc way of sorting the data in an order so that observations are close in terms of the row they occupy if they have similar values of (n_1, n_2, n_3).

2. We sort the Monte Carlo dataset in ascending order according to:

$$v_1 + v_4 + |v_2 + v_5| + \sin(2\pi Q(v_3 + v_6)) \tag{G.1}$$

where the $Q(x)$ function maps every entry of a vector to the unit interval based on its ordinal relationship with other entries of the vector[10] Note that we do not use v_7 or v_8 in this function - these are irrelevant variables that ideally the Monte Carlo will learn to

[9]One possibility is to use reinforcement learning to learn optimal behaviour. However, this would involve a complex reinforcement learning process and we would have the problem of testing a complex model using a Monte Carlo that depends on another complex model. We would then need to test the validity of the reinforcement learning process. Rather than do this we have decided on a simpler approach.

[10]As an example if $x = [4, 2, 3, 1, 0]$ then $Q(x) = [1, 0.5, 0.75, 0.25, 0]$.

ignore.

3. After both of the above sorts, the row i of the Monte Carlo dataset is associated with the row i of the real dataset and gets its (n_1, n_2, n_3) tuple

There is no deep logic that lead us to equation G.1. However, it ensures a nonlinear and complex relationship between the values of $\{v_k\}_{k=1:8}$ and the row that an observation occupies. Combined with sorting the real dataset, this ensures that there is a weak relationship between the values of $\{v_k\}_{k=1:8}$ and the $(n_1, n_2, n_3, p, \tilde{p})$ tuple. This procedure results in a relationship between $\{v_k\}_{k=1:8}$ and expected demand that is slightly weaker (in terms of R^2) than what we get obtain in our real data case. This is ideal for testing the ability of our neural network method to be able to estimate costs effectively. The lower R^2 implies that our method will perform slightly better on the real data than in this Monte Carlo scenario.

It is worth noting that this data generation process also implies that $\{v_k\}_{k=1:8}$ could be used to predict p and $\tilde{p}$ since these variables were part of the demand estimation used to sort the real dataset in the initial step above. We believe that this is realistic however as some of the inputs to our demand function may indeed by correlated to price and preceding price. For instance, if the standard deviation of bids in the second round is low that might imply prices in the third round are high as undercutting has not taken place.

Given our dataset with columns for $\{v_k\}_{k=1:8}, n_1, n_2, n_3, p, \tilde{p}$ we also calculate the implied cost with equation 5 and the demand with equation 6. We then use the demand probability to randomly generate a binary variable y that describes whether the bidder won the auction. We then apply our procedure to estimate y given only $\{v_k\}_{k=1:8}, p, \tilde{p}$ as inputs. Once we have a trained procedure we get new estimates of $\hat{n}_1, \hat{n}_2, \hat{n}_3$ from the trained model and use these to estimate cost for each observation.

We begin by showing the convergence of each Monte Carlo experiment in Figure G.1. The final convergence achieved is summarized in Table G.1. The accuracy of the probability of winning function is around 3.5%, which is less than the approximate 6.5% that can be seen for the benchmark specification in table J.1.2. This suggests that the error we see in our Monte Carlo case is likely to be more extreme than what occurs in the real data case.

The differences between the real and predicted costs from our Monte Carlo experiments is summarized in table G.2. We can see that the difference between actual and predicted cost is close to zero with a standard deviation around 10% of the expected price.

There are a few limitations of this Monte Carlo analysis. First, our Monte Carlo assumes that our demand function (6) is correctly specified. Second, we do not model the possibility of bidders in our

Table (G.1) Final Convergence of Models Trained With Monte Carlo Data

			R2			Pearson Corr			MAE		
IW	Seed	Num. Iterates	All	Test	Training	All	Test	Training	All	Test	Training
0.0	1	270	0.0413	0.0415	0.0413	0.2048	0.2051	0.2048	0.2486	0.2488	0.2486
0.0	2	326	0.0222	0.0246	0.0219	0.1518	0.1587	0.151	0.2542	0.2535	0.2543
0.0	3	440	0.0358	0.0341	0.036	0.1981	0.194	0.1986	0.2421	0.2428	0.242
0.0	4	818	0.0303	0.0296	0.0304	0.176	0.1735	0.1763	0.251	0.2502	0.2511
0.0	5	1174	0.0435	0.0451	0.0434	0.2101	0.2132	0.2098	0.2478	0.2468	0.2479
0.5	1	695	0.0399	0.0396	0.04	0.2003	0.1995	0.2004	0.2526	0.2529	0.2526
0.5	2	1077	0.0352	0.0363	0.0351	0.1897	0.1925	0.1894	0.2492	0.2487	0.2493
0.5	3	742	0.0304	0.0286	0.0306	0.1862	0.182	0.1867	0.2423	0.2429	0.2422
0.5	4	912	0.0354	0.0346	0.0355	0.1891	0.1868	0.1893	0.2522	0.2515	0.2523
0.5	5	1440	0.0437	0.0446	0.0436	0.2105	0.2122	0.2103	0.2474	0.2466	0.2475
1.0	1	297	0.0283	0.0277	0.0284	0.1696	0.1679	0.1698	0.2594	0.2597	0.2593
1.0	2	738	0.0376	0.0398	0.0374	0.195	0.2002	0.1944	0.2512	0.2505	0.2513
1.0	3	392	0.0367	0.0342	0.037	0.1938	0.1883	0.1944	0.2492	0.2499	0.2491
1.0	4	1232	0.0394	0.0404	0.0393	0.1995	0.2015	0.1993	0.251	0.2491	0.2512
1.0	5	848	0.0382	0.0388	0.0381	0.1961	0.1976	0.1959	0.2526	0.2517	0.2527

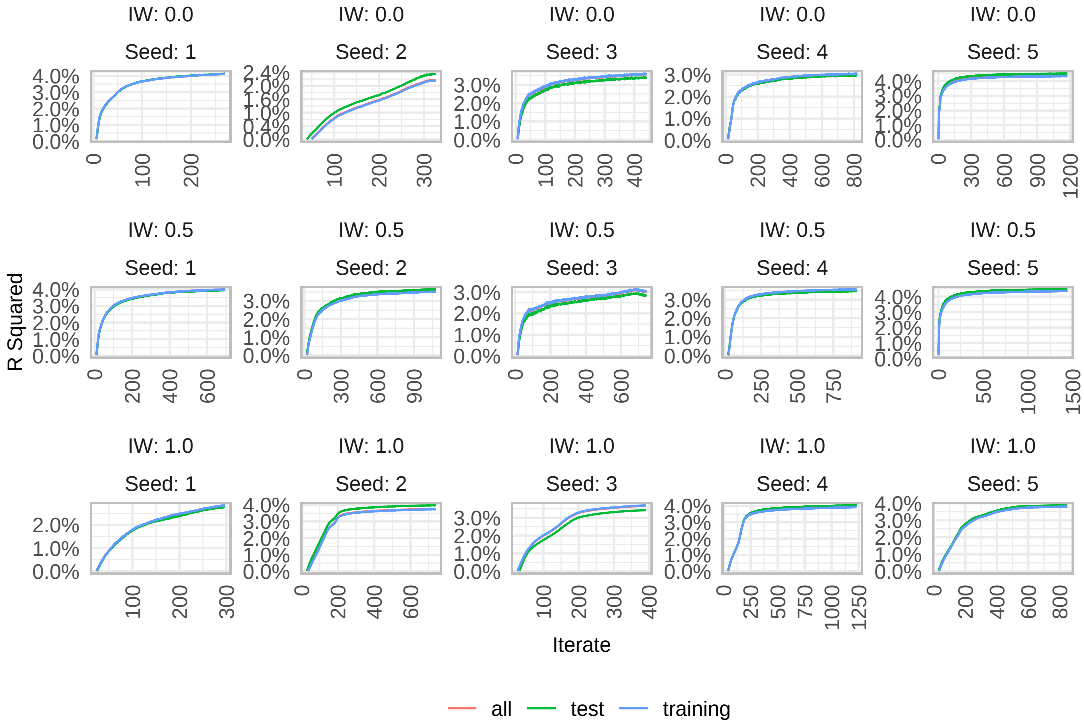
Note: This table shows the R^2, Pearson correlation and Mean Absolute Error (MAE) of the prediction at the point when training stops. Due to the early stopping logic, training stops after 50 iterates in which the loss function on the test set has not improved. IW refers to the weight given to an identity covariance matrix with $1 - \text{IW}$ being the weight towards a random covariance matrix drawn from the inverse Wishart distribution.

Table (G.2) Differences Between Predicted and Actual Costs in Several Monte Carlo Experiments

		Cor.$(\frac{c}{p}, \frac{\hat{c}}{p})$		Mean Cost		Mean Gap	St.Dev Gap
Identity Mix	Seed	Pearson	Spearman	Actual	Predicted	$c - \hat{c}$	$c - \hat{c}$
0.0	1	0.9037	0.9004	0.6287	0.5955	0.0331	0.0406
0.0	2	0.7255	0.6893	0.6326	0.5955	0.0371	0.0632
0.0	3	0.9085	0.8946	0.6152	0.5955	0.0197	0.0405
0.0	4	0.8234	0.8065	0.5851	0.5955	-0.0104	0.0549
0.0	5	0.8696	0.8486	0.5701	0.5955	-0.0254	0.0492
0.5	1	0.9134	0.9056	0.5836	0.5955	-0.012	0.0393
0.5	2	0.8591	0.8407	0.5787	0.5955	-0.0168	0.0499
0.5	3	0.9036	0.8892	0.5965	0.5955	9e-04	0.0407
0.5	4	0.858	0.8445	0.5766	0.5955	-0.019	0.0505
0.5	5	0.8843	0.8767	0.5644	0.5955	-0.0311	0.0468
1.0	1	0.8783	0.8672	0.6153	0.5955	0.0198	0.0441
1.0	2	0.9019	0.8999	0.5812	0.5955	-0.0144	0.0436
1.0	3	0.868	0.8494	0.6196	0.5955	0.0241	0.0464
1.0	4	0.8865	0.879	0.5699	0.5955	-0.0257	0.0466
1.0	5	0.8998	0.8973	0.57	0.5955	-0.0255	0.0442

Note: This table shows the pearson correlation and spearman correlation between actual and predicted costs when both are divided by the price. Correlations are higher when costs are not divided by price. The mean cost, mean predicted costs, mean gap between these costs and the standard deviations of these gaps are also presented. The seed in the first column is used for all randomisation used in constructing the Monte Carlo dataset.

Figure (G.1) R^2 at Each Iteration Achieved by Each Monte Carlo Experiment



Note: This figure shows the R^2 values achieved at each training iteration of each Monte Carlo experiment. Data is only shown for positive R^2 values, hence a line will generally not appear until several training iterations have been completed. Early stopping logic causes training to end at different iterations for each case. Note that each Monte Carlo experiment used a completely different dataset and test set. Due to randomness and training methods aimed at minimizing overfitting, test set error was occasionally lower than training error. IW refers to the weight given to an identity covariance matrix with $1 - \mathrm{IW}$ being the weight towards a random covariance matrix drawn from the inverse Wishart distribution.

dataset not bidding with optimal behaviour, an assumption that underlies the derivation of equation 5. However, our Monte Carlo does test the ability of our framework to accurately estimate costs, assuming the optimality of bidder decisions and that our assumed demand function is accurate.

Table (H.1) Results of alternative estimations

	Training			Test		
Model	R^2	Pearson	Spearman	R^2	Pearson	Spearman
Unconstrained Neural Network	8.20%	0.292	0.270	8.12%	0.291	0.268
Random Forest	7.96%	0.307	0.282	7.94%	0.307	0.281
Regression Tree	7.09%	0.266	0.208	7.04%	0.265	0.209
Probit	3.56%	0.190	0.194	3.60%	0.191	0.196

H Accuracy of Alternative Demand Estimation Models

We use alternative machine learning methods to predict demand as a contrast with the accuracy of the neural network method presented in section 4.1. Specifically, we have done this for a probit regression, a random forest, a regression tree[11] as well as an unconstrained neural network. In these cases, we simply input all ten predictive variables (the eight variables that enter our benchmark neural network as well as price and the difference between price and the best preceding price)[12] to the respective method. The probit regression is done with Julia's **GLM** package.[13] The random forest is done with Julia's **DecisionTree** package and includes 500 trees, 3 features being selected at each node, 50% of the data being selected for each tree and a maximum tree depth of 3. The regression tree (also done with **DecisionTree**) does not use randomisation. It selects from any features at every node with a maximum depth of 3. The unconstrained neural network has the aforementioned ten inputs, six hidden layers each of which has four leaky ReLU neurons and a single output that is passed through a sigmoid function to return a probability. It is trained using the loss function in equation 10 with the same early stopping logic as the benchmark specification.

We repeat all estimations ten times and report the median R^2, Pearson and Spearman correlations in Table H.1. Contrasting this with our benchmark specifications performance (in Table J.1.2), it can be seen that our benchmark estimates performed with an R^2 around 6.5% and correlations (Pearson & Spearman) between predictions and realised demand of around 25.5%. This performance is slightly worse than that arrived at through a random forest, regression tree or an unconstrained neural network, however, it is significantly better than a probit regression. The outperformance of the unconstrained neural network and random forest are likely due to them having a high degree of flexibility. Our benchmark specification however is constrained to have demand decreasing as price increases at a relatively smooth gradient. As a result of this lower flexibility, it cannot achieve the same fit as the more flexible models. Finally, the probit model is highly constrained and achieves significantly worse prediction accuracy than the other techniques.

[11] Our problem is one of classification but for comparison with other models we have a tree that outputs the fraction of training set observations at a particular leaf that did make a sale, a fraction that can be interpreted as a probability. We do not assign a label but just use this probability.

[12] While the best preceding price $\tilde{p}$ is used in our benchmark method, we instead use $p - \tilde{p}$ in this case. While this is redundant in the probit case, it does make it slightly easier to train in the random forest and neural networks case.

[13] We also perform a probit Lasso estimation using Julia's **Lasso** package. We found that the best coefficient penalty term (in terms of delivering the highest R^2 on the test set) was usually zero. Hence we do not present these results as they are the same as the basic probit estimation.

I Parallels with the Guerre, Perrigne and Vuong (2000) method

A benchmark paper in the empirical auctions literature is that of Guerre, Perrigne and Vuong (2000). This methodology has also been reviewed in the context of the broader empirical auction literature by Perrigne and Vuong (2019, section 2.1). In the following discussion, we refer to (Perrigne and Vuong, 2019) presentation of the method.

These papers differ from our own most noticeably by considering bidders with private valuations buying a good. On the other hand, we consider bidders with private costs selling a good/service. The more fundamental difference however is in the methodology by which we estimate private costs as compared to how Guerre, Perrigne and Vuong (2000) estimate private valuations. Specifically our paper uses machine learning to approximate the demand function. On the other hand Guerre, Perrigne and Vuong (2000) focuses on first price sealed bid auctions and is able to derive many expressions using game theoretic analysis.

To more precisely describe the point at which these two methods differ, we summarize the logical steps of each approach in Figure I.1. Our method takes the path on the left of the page while the method of Guerre, Perrigne and Vuong (2000) takes the path on the right. Note that, rather than adhering strictly to the presentation in Perrigne and Vuong (2019, section 2.1), we have made an adjustment to focus on cost estimation for sellers instead of valuation estimation for buyers. This adjustment is done so that the two paths are more easily comparable.

The notation used is as follows: p is price; c is the cost that we seek to infer; $D(p)$ is the probability of winning function; $D'(p)$ is the derivative of the probability of winning function; $F(\cdot)$ is the cdf of the distribution of costs; I is the number of bidders.

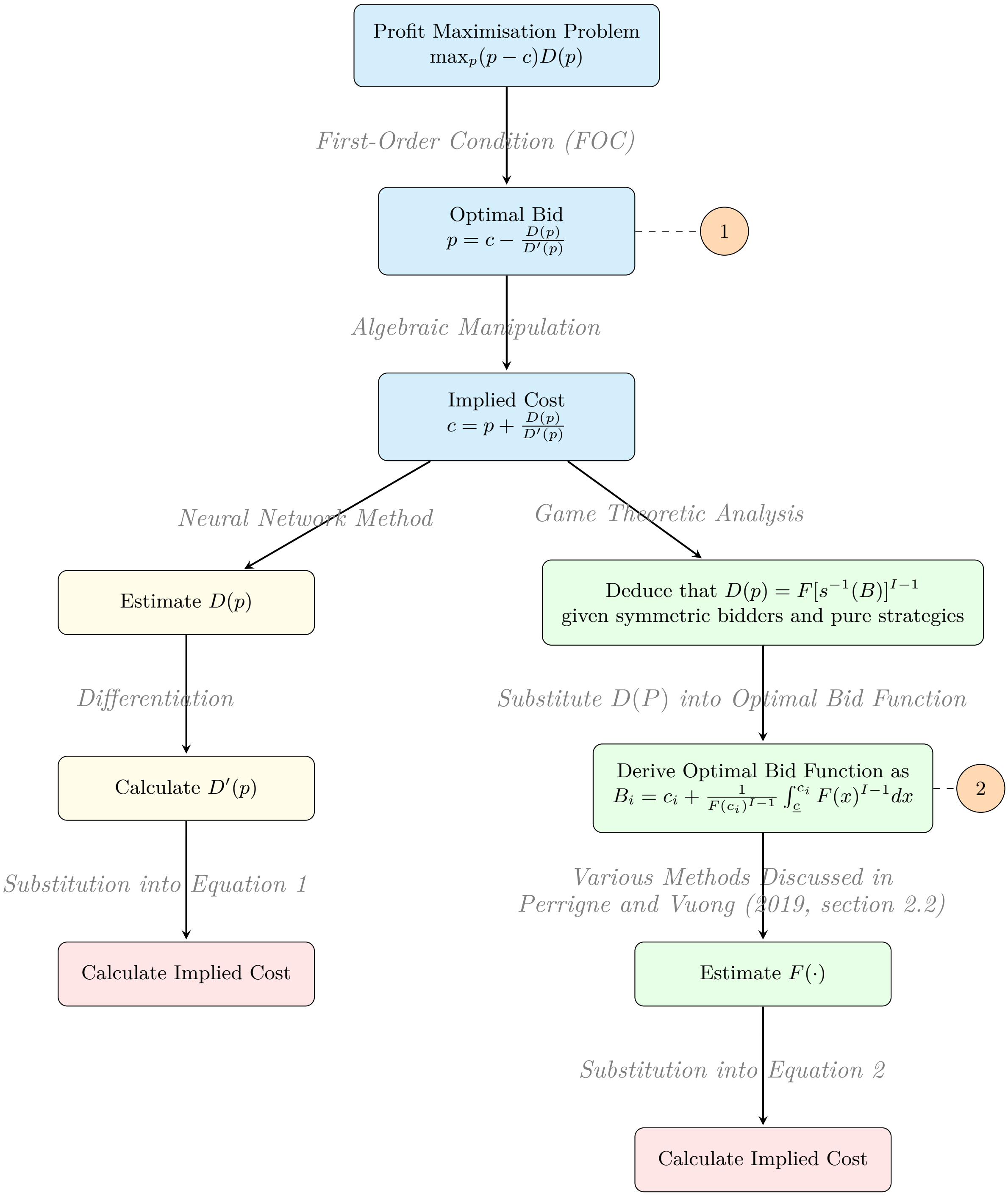


Figure (I.1) Our methodology (left path) and the benchmark by Guerre, Perrigne and Vuong (2000) as discussed in Perrigne and Vuong (2019) (right path) both start by deriving the first-order condition from the bidder's profit maximization problem. However, due to the complexity of our auction format, we cannot proceed further using traditional methods. In contrast, Guerre, Perrigne and Vuong (2000) analyze the simpler first-price sealed-bid auction, allowing them to make more progress through game-theoretic analysis. As a result, the two methods become different after the first order condition stage.

J Alternative cost model specifications

Several choices need to be made in designing the specific neural network outlined in section 4. Nonetheless, our results tend to remain robust to other sensible choices. To demonstrate this, we present key tables and figures derived from following the same methodology as section 4, albeit with deviations in certain modelling choices. Table J.1 offers a summary of these alternative specifications.

Table (J.1) Guide to alternative specifications used

Specification	Loss function	Activation function	Cost floor	Dropout Proportion	Num. hidden layers	Special Training Settings
J.1	ML & MSE	Leaky ReLU	25%	2.5%	6	None
J.2	MSE	Leaky ReLU	25%	2.5%	6	None
J.3	ML & MSE	Tanh	25%	2.5%	6	None
J.4	ML & MSE	Swish	25%	2.5%	6	None
J.5	ML & MSE	Leaky ReLU, Tanh and Swish	25%	2.5%	6	None
J.6	ML & MSE	Leaky ReLU	0%	2.5%	6	None
J.7	ML & MSE	Leaky ReLU	20%	2.5%	6	None
J.8	ML & MSE	Leaky ReLU	30%	2.5%	6	None
J.9	ML & MSE	Leaky ReLU	40%	2.5%	6	None
J.10	ML & MSE	Leaky ReLU	50%	2.5%	6	None
J.11	ML & MSE	Leaky ReLU	62.5%	2.5%	6	None
J.12	ML & MSE	Leaky ReLU	75%	2.5%	6	None
J.13	ML & MSE	Leaky ReLU	25%	0%	6	None
J.14	ML & MSE	Leaky ReLU	25%	5%	6	None
J.15	ML & MSE	Leaky ReLU	25%	2.5%	3	None
J.16	ML & MSE	Leaky ReLU	25%	2.5%	6	Trained with last bidder obs
J.17	MSE	Tanh and Swish	20%	4%	5	None
J.18	ML & MSE	Leaky ReLU	25%	2.5%	6	Trained with fewer input vars
J.19	ML & MSE	Leaky ReLU	25%	2.5%	6	Trained with more input vars
J.20	ML & MSE	Leaky ReLU	25%	2.5%	6	Hidden layers of width 2
J.21	ML & MSE	Leaky ReLU	25%	2.5%	6	Hidden layers of width 3

Note: This table presents the 21 alternative specifications we used. The first column numbers the specifications and provides a section reference. Column (2) specifies the loss functions: Maximum Likelihood (ML) and Mean Squared Error (MSE). Activation functions (column 3) vary across specifications: Leaky Rectified Linear Unit (Leaky ReLU), hyperbolic tangent (Tanh), and Swish. The cost floor (column 4) represents the minimum value for the cost, expressed as a percentage of the final bid. The dropout proportion (column 5) indicates the fraction of input units to drop during training. The number of hidden layers in the neural network is specified in column (6). We note special training settings in the last column.

J.1 Specification 1

This is the benchmark specification. Some tables and figures are repeated from those in section 4 for easy comparability to the full set of other figures and tables that are included in this appendix. We also replicate Table 4 for winners in Table J.1.1, where the last month is defined as a dummy variable that equals one if the auction is finalised in December and zero otherwise.

Table (J.1.1) The Year-End Effect for Margins

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
last_mont_dec	4.841***	3.020***	1.591***	3.161***	1.646***	2.184***	1.108*	1.157*
	(0.582)	(0.581)	(0.585)	(0.592)	(0.584)	(0.592)	(0.635)	(0.634)
R2	0.002	0.031	0.106	0.031	0.108	0.055	0.116	0.119
N	136756	136751	123531	132021	123524	132016	118433	118424
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	No	Yes	No	No	Yes	Yes	No	Yes
Seller FE	No	No	Yes	No	Yes	No	Yes	Yes
Buyer FE	No	No	No	Yes	No	Yes	Yes	Yes

Note: The sample is composed of winners of competitive lots for which we estimated costs. The dependent variable in all columns is a margin common cost as defined in equation 12. Last month is a dummy variable that takes the value one if the tender was finalised in December and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and sector FE are sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and by seller.

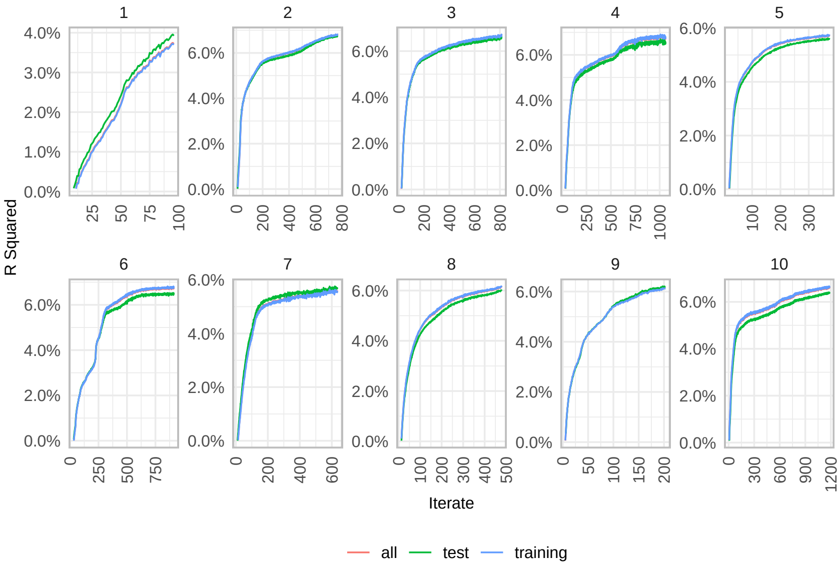
Table (J.1.2) Final convergence

		R^2			Pearson Corr			MAE		
Seed	Num. Iterates	All	Test	Training	All	Test	Training	All	Test	Training
1	96	0.0372	0.0393	0.037	0.1949	0.2006	0.1943	0.2575	0.2574	0.2575
2	769	0.068	0.0676	0.0681	0.2621	0.2617	0.2622	0.252	0.2532	0.2519
3	814	0.0668	0.0659	0.0669	0.2603	0.2589	0.2605	0.2517	0.2521	0.2516
4	1060	0.068	0.0659	0.0682	0.2618	0.2576	0.2622	0.2542	0.2545	0.2542
5	373	0.0575	0.0563	0.0577	0.24	0.2374	0.2403	0.2595	0.2596	0.2595
6	912	0.0676	0.0654	0.0678	0.26	0.256	0.2604	0.2561	0.2555	0.2561
7	637	0.0557	0.0568	0.0556	0.2489	0.2518	0.2486	0.2449	0.2442	0.245
8	479	0.0613	0.0599	0.0614	0.2491	0.2464	0.2494	0.252	0.253	0.2519
9	203	0.0613	0.062	0.0613	0.248	0.2492	0.2479	0.2538	0.253	0.2539
10	1183	0.0658	0.0635	0.066	0.2577	0.2526	0.2582	0.2523	0.2517	0.2524

Note: This table shows the R^2, Pearson correlation and Mean Absolute Error (MAE) of the prediction at the point when training stops. Due to the early stopping logic, training stops after 50 iterates in which performance on the loss function on the test set has not improved.

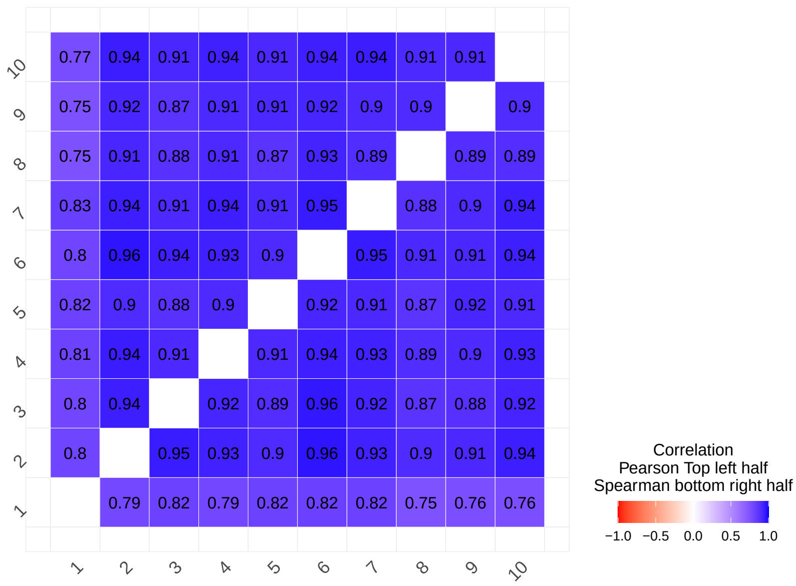
We first present tables and figures summarising the convergence path and fit of the model. The convergence plot of each seed with this specification can be seen in Figure J.1.1, while the correlation matrix between implied costs generated with each seed is displayed in Figure J.1.2. We summarise the final convergences achieved by each seed in Table J.1.2. The fraction of demands and costs that fall outside the feasible range is in Table J.1.3. Note that models that do not converge well may have more estimated probabilities outside the unit interval. Our analysis only uses the 5 trained models that have

Figure (J.1.1) Convergence Plot for Specification 1



Note: This chart plots the convergence of each trained model in terms of the R^2. We only plot positive R^2 for each which is why some seeds appear to start later on. In general, R^2 values start training at negative values.

Figure (J.1.2) Correlation Matrix Between Costs Implied by Each Seed for Specification 1



Note: This chart plots the correlation between $\frac{\text{Implied cost}}{\text{final price bid}}$ using the implied costs from each trained model. The top left half of the matrix are Pearson correlations while the bottom right half are Spearman correlations.

Table (J.1.3) Proportion of invalid values for each trained model.

Seed	Demand > 1	Demand < 0	Cost > Bid	Cost < 0
Combined	0.0002%	0.0002%	0.0537%	0
1	0%	0%	0%	0
2	0.0004%	0.0004%	0.058%	0
3	0.0001%	0.0001%	0.2064%	0
4	0.0041%	0.0041%	0%	0
5	0.0001%	0.0001%	0.0009%	0
6	0.0037%	0.0037%	0.3208%	0
7	0%	0%	0.022%	0
8	0.0006%	0.0006%	0.2715%	0
9	0.0004%	0.0004%	0.7095%	0
10	0.0005%	0.0005%	0%	0

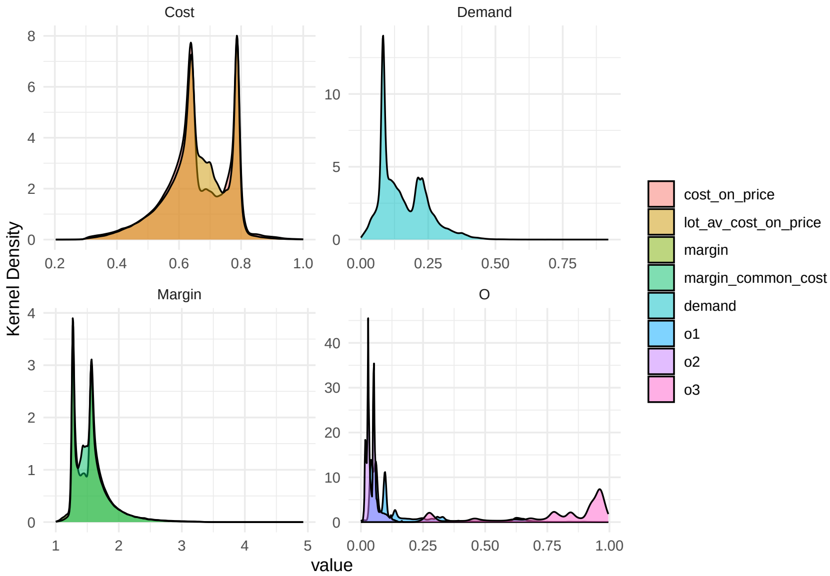
Note: This table shows fraction of demands that are outside the unit interval and the fraction of costs that are above the final bid cost or below zero. Each figure in this table is a percentage of the non-missing values in the dataset. Figures are expressed to 4 decimal places.

Table (J.1.4) Variable Importance Plot

Permuted Variable	Relative MSE	Relative Spearman Correlation
Price	11.1979	0.6656
Best Preceding Price	29.3061	0.716
Log Expected Price	1.0058	0.9585
Num. Bidders after	1.0581	0.5045
Mean prices bidder	1.0747	0.5535
Num. Participants	1.0387	0.7754
Prop. Inactive Bidders	1.0386	0.7427
Price Reduction in Round 3	1.0021	0.9863
Round 2 std of bids	1.136	0.4402
Year tender published	1.0025	0.9823

Note: To construct this table we took the 5 trained models (of 10 in total) with above median performance on the test set. We recorded it's performance on the full (test + training) sample as a benchmark. We then permuted each of the input variables (30 times per model) and recorded their performance. We recorded the relative MSE and Spearman correlation for each permutation and each model. We then took the medians for each model. Finally we took the median across models to get one relative performance metric per permuted variable. A relative MSE above 1 indicates a deterioration in performance by permuting that variable while a relative spearman below 1 indicates the same.

Figure (J.1.3) Kernel Densities of Values Estimated with Specification 1



Note: This chart plots the kernel densities of demand, implied margin, cost and the o_1, o_2, o_3 parameters. For the purposes of this figure, we remove observations where $\frac{\text{Implied cost}}{\text{final price bid}}$ is outside the $[0, 1]$ range or demand is outside this same range. The amount of observations that are dropped in this way are small and represented in Table J.1.3.

the highest R^2 however these poorly performing cases are not used. The variable importance of each of the model's input variables is summarised in Table J.1.4. We then present kernel densities of demand, cost, margin, o_1, o_2 and o_3 in Figure J.1.3. In Table J.1.5 (corresponding to Table 4) we provide the regression coefficients of margin with different fixed effects for winners and losers.

The regressions with the interaction terms are in Table J.1.6 (which corresponds to Table 5).

Finally in Table J.1.6 we replicate the regressions without controls on different time intervals of data. Specifically, for all bids, winners and losers respectively we perform separate regressions: one on 2017-2019 (pre-Covid period) and one on 2020-2021 (where the bidding process was impacted by covid). We present our findings in Table J.1.7.

Table (J.1.5) The Year-End Effect for Margins (with Varying Fixed Effects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Winners								
Last month	2.34***	2.3***	1.88***	1.77***	1.97***	1.88***	1.77***	1.86***
	(0.424)	(0.414)	(0.402)	(0.379)	(0.402)	(0.378)	(0.399)	(0.399)
R^2	0.00146	0.0319	0.3	0.142	0.302	0.164	0.403	0.405
N	136756	136751	136756	136756	136751	136751	136756	136751
Panel B: Losers								
Last month	3.62***	3.39***	2.94***	3.0***	3.04***	2.96***	2.76***	2.85***
	(0.558)	(0.541)	(0.488)	(0.479)	(0.488)	(0.475)	(0.458)	(0.457)
R^2	0.002	0.0375	0.232	0.126	0.235	0.151	0.32	0.323
N	382559	382550	382559	382559	382550	382550	382559	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	No	Yes	No	No	Yes	Yes	No	Yes
Seller FE	No	No	Yes	No	Yes	No	Yes	Yes
Buyer FE	No	No	No	Yes	No	Yes	Yes	Yes

Note: The sample is composed of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is a margin common cost as defined in equation 12. Last month is a dummy variable that takes the value one if the tender was finalised in December or November and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and by seller. *p<0.1; **p<0.05; ***p<0.01.

Table (J.1.6) The Year-End Effect for Margins (with Varying Controls)

	All bids			*Winners*			*Losers*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Last month	2.41***	1.73***	1.61***	1.86***	1.2***	1.48***	2.85***	2.09***	1.82***
	(0.402)	(0.379)	(0.455)	(0.399)	(0.384)	(0.507)	(0.457)	(0.429)	(0.52)
Last month × 4 bidders			0.647			0.127			0.898
			(0.638)			(0.764)			(0.684)
Last month × 5 bidders			1.26			0.201			1.6*
			(0.859)			(1.07)			(0.906)
Last month × 6 bidders			-1.14			-2.18*			-0.77
			(1.05)			(1.27)			(1.12)
Last month × 7+ bidders			-1.36			-2.24*			-1.34
			(1.0)			(1.19)			(1.07)
4 bidders		-0.676***	-0.777***		-2.91***	-2.93***		-0.428	-0.571**
		(0.237)	(0.254)		(0.265)	(0.288)		(0.262)	(0.281)
5 bidders		-4.17***	-4.36***		-7.85***	-7.88***		-4.07***	-4.31***
		(0.31)	(0.32)		(0.36)	(0.384)		(0.33)	(0.344)
6 bidders		-7.7***	-7.55***		-12.0***	-11.7***		-7.66***	-7.56***
		(0.387)	(0.412)		(0.465)	(0.497)		(0.415)	(0.439)
7+ bidders		-13.5***	-13.4***		-15.4***	-15.1***		-14.3***	-14.1***
		(0.412)	(0.427)		(0.499)	(0.509)		(0.441)	(0.456)
R^2	0.291	0.301	0.301	0.405	0.418	0.418	0.323	0.334	0.334
N	519301	519301	519301	136751	136751	136751	382550	382550	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Columns 1, 4, and 7 correspond to specification 13. Columns 2, 5, and 8 also include dummies for each number of bidders in the auction. Columns 3, 6, and 9 further split the last month dummy by the number of bidders in the auction. Year FE stands for year fixed effects, buyer FE refers to government department fixed effects, seller FE indicates firm fixed effects, and sector FE represents sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

Table (J.1.7) The Year-End Effect for Margins (with Split by Year Interval)

	All 17-19	*All 20-21*	*Winners 17-19*	*Winners 20-21*	*Losers 17-19*	*Losers 20-21*
	(1)	(2)	(3)	(4)	(5)	(6)
Last month	2.33***	2.16***	2.13***	1.41***	2.62***	2.54***
	(0.511)	(0.525)	(0.611)	(0.543)	(0.592)	(0.598)
R^2	0.374	0.342	0.501	0.489	0.416	0.377
N	208764	310537	56309	80442	152455	230095
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation 12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

J.2 Specification 2

This specification is the same as the benchmark specification but with the loss function being Mean Squared Error (MSE) rather than the ML and MSE combination approach in equation 10.

Table (J.2.1) Final convergence

		R^2			Pearson Corr			MAE		
Seed	Num. Iterates	All	Test	Training	All	Test	Training	All	Test	Training
1	680	0.0558	0.0587	0.0555	0.2392	0.2464	0.2384	0.2545	0.2542	0.2545
2	904	0.0689	0.0682	0.069	0.2638	0.2627	0.2639	0.2523	0.2535	0.2522
3	1038	0.0679	0.0669	0.068	0.2626	0.2611	0.2628	0.2513	0.2518	0.2513
4	960	0.0676	0.0653	0.0678	0.2606	0.256	0.2611	0.2548	0.2551	0.2548
5	840	0.0612	0.0603	0.0613	0.2481	0.2463	0.2483	0.2611	0.2611	0.2611
6	719	0.067	0.0647	0.0673	0.2589	0.2546	0.2594	0.2561	0.2556	0.2561
7	377	0.0537	0.0551	0.0535	0.2384	0.2419	0.2381	0.2485	0.2478	0.2486
8	1157	0.0685	0.068	0.0686	0.2648	0.2637	0.2649	0.2536	0.2545	0.2535
9	1023	0.0685	0.0694	0.0685	0.2623	0.2637	0.2621	0.254	0.2532	0.2541
10	1371	0.0662	0.0641	0.0664	0.2583	0.2538	0.2589	0.2527	0.2522	0.2528

Note: This table shows the R^2, Pearson correlation and Mean Absolute Error (MAE) of the prediction at the point when training stops. Due to the early stopping logic, training stops after 50 iterates in which performance on the loss function on the test set has not improved.

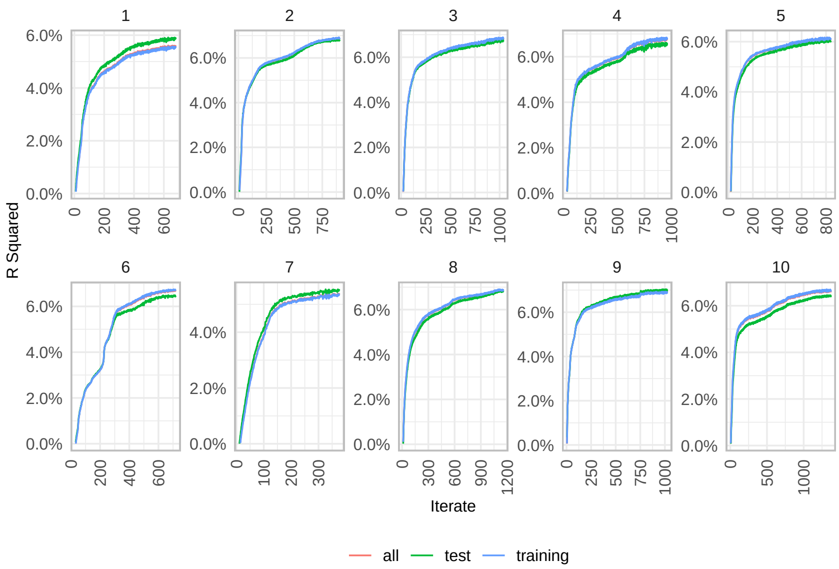
Table (J.2.2) Proportion of invalid values for each trained model.

Seed	Demand > 1	Demand < 0	Cost > Bid	Cost < 0
Combined	0.0004%	0.0004%	0.0615%	0
1	0%	0%	0%	0
2	0.0006%	0.0006%	0.0791%	0
3	0.0002%	0.0002%	0.2044%	0
4	0.0044%	0.0044%	0%	0
5	0.0001%	0.0001%	0%	0
6	0.0031%	0.0031%	0.3662%	0
7	0%	0%	0.0521%	0
8	0.0004%	0.0004%	0.0783%	0
9	0.0007%	0.0007%	0.0082%	0
10	0.0006%	0.0006%	0%	0

Note: This table shows fraction of demands that are outside the unit interval and the fraction of costs that are above the final bid cost or below zero. Each figure in this table is a percentage of the non-missing values in the dataset. Figures are expressed to 4 decimal places.

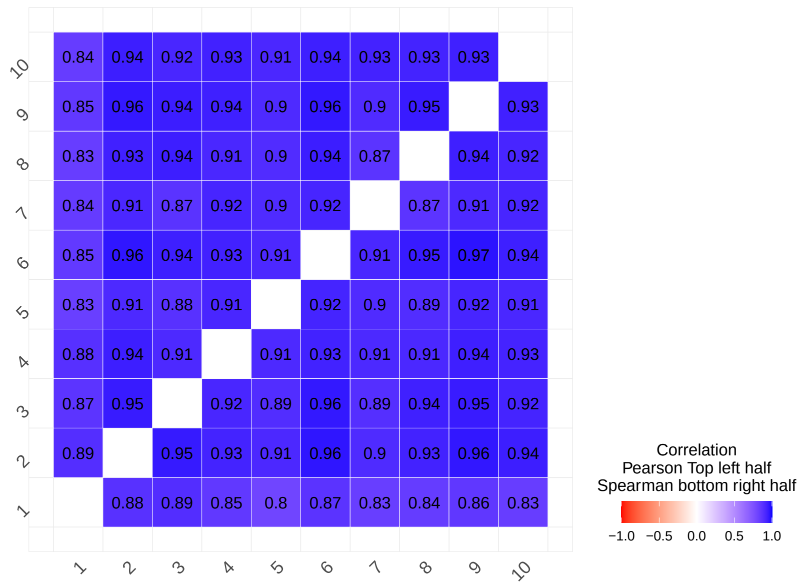
We first present tables and figures summarising the convergence path and fit of the model. The convergence plot of each seed with this specification can be seen in Figure J.2.1, while the correlation matrix between implied costs generated with each seed is displayed in Figure J.2.2. We summarise the final convergences achieved by each seed in Table J.2.1. The fraction of demands and costs that fall outside the feasible range is in Table J.2.2. Note that models that do not converge well may have more estimated probabilities outside the unit interval. Our analysis only uses the 5 trained models that have the highest R^2 however these poorly performing cases are not used. The variable importance of each of

Figure (J.2.1) Convergence Plot for Specification 2



Note: This chart plots the convergence of each trained model in terms of the R^2. We only plot positive R^2 for each which is why some seeds appear to start later on. In general, R^2 values start training at negative values.

Figure (J.2.2) Correlation Matrix Between Costs Implied by Each Seed for Specification 2



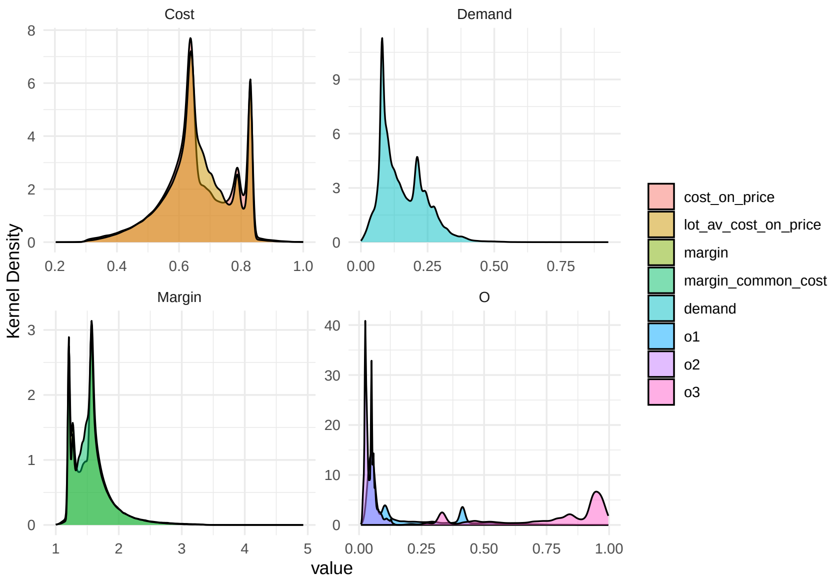
Note: This chart plots the correlation between $\frac{\text{Implied cost}}{\text{final price bid}}$ using the implied costs from each trained model. The top left half of the matrix are Pearson correlations while the bottom right half are Spearman correlations.

Table (J.2.3) Variable Importance Plot

Permuted Variable	Relative MSE	Relative Spearman Correlation
Price	11.1197	0.6559
Best Preceding Price	26.7649	0.7171
Log Expected Price	1.0058	0.962
Num. Bidders after	1.0591	0.4874
Mean prices bidder	1.0752	0.5394
Num. Participants	1.0414	0.7581
Prop. Inactive Bidders	1.0437	0.7418
Price Reduction in Round 3	1.0019	0.9907
Round 2 std of bids	1.1285	0.449
Year tender published	1.0025	0.9827

Note: To construct this table we took the 5 trained models (of 10 in total) with above median performance on the test set. We recorded it's performance on the full (test + training) sample as a benchmark. We then permuted each of the input variables (30 times per model) and recorded their performance. We recorded the relative MSE and Spearman correlation for each permutation and each model. We then took the medians for each model. Finally we took the median across models to get one relative performance metric per permuted variable. A relative MSE above 1 indicates a deterioration in performance by permuting that variable while a relative spearman below 1 indicates the same.

Figure (J.2.3) Kernel Densities of Values Estimated with Specification 2



Note: This chart plots the kernel densities of demand, implied margin, cost and the o_1, o_2, o_3 parameters. For the purposes of this figure, we remove observations where $\frac{\text{Implied cost}}{\text{final price bid}}$ is outside the $[0, 1]$ range or demand is outside this same range. The amount of observations that are dropped in this way are small and represented in Table J.2.2.

the model's input variables is summarised in Table J.2.3. We then present kernel densities of demand, cost, margin, o_1, o_2 and o_3 in Figure J.2.3. In Table J.2.4 (corresponding to Table 4) we provide the regression coefficients of margin with different fixed effects for winners and losers.

Table (J.2.4) The Year-End Effect for Margins (with Varying Fixed Effects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Winners								
Last month	2.19***	2.13***	1.67***	1.62***	1.76***	1.73***	1.58***	1.67***
	(0.421)	(0.41)	(0.399)	(0.376)	(0.398)	(0.374)	(0.396)	(0.396)
R^2	0.00142	0.0317	0.301	0.141	0.303	0.163	0.404	0.406
N	136756	136751	136756	136756	136751	136751	136756	136751
Panel B: Losers								
Last month	3.43***	3.17***	2.68***	2.82***	2.78***	2.77***	2.52***	2.61***
	(0.555)	(0.535)	(0.481)	(0.476)	(0.481)	(0.471)	(0.454)	(0.453)
R^2	0.00229	0.0375	0.232	0.125	0.236	0.15	0.32	0.323
N	382559	382550	382559	382559	382550	382550	382559	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	No	Yes	No	No	Yes	Yes	No	Yes
Seller FE	No	No	Yes	No	Yes	No	Yes	Yes
Buyer FE	No	No	No	Yes	No	Yes	Yes	Yes

Note: The sample is composed of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is a margin common cost as defined in equation 12. Last month is a dummy variable that takes the value one if the tender was finalised in December or November and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and by seller. *p<0.1; **p<0.05; ***p<0.01.

The regressions with the interaction terms are in Table J.2.5 (which corresponds to Table 5).

Finally in Table J.2.5 we replicate the regressions without controls on different time intervals of data. Specifically, for all bids, winners and losers respectively we perform separate regressions: one on 2017-2019 (pre-Covid period) and one on 2020-2021 (where the bidding process was impacted by covid). We present our findings in Table J.2.6.

Table (J.2.5) The Year-End Effect for Margins (with Varying Controls)

	All bids			*Winners*			*Losers*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Last month	2.19***	1.63***	1.55***	1.67***	1.1***	1.4***	2.61***	1.98***	1.76***
	(0.398)	(0.379)	(0.455)	(0.396)	(0.383)	(0.508)	(0.453)	(0.429)	(0.522)
Last month × 4 bidders			0.588			0.0524			0.847
			(0.638)			(0.763)			(0.684)
Last month × 5 bidders			1.15			0.155			1.48
			(0.857)			(1.07)			(0.903)
Last month × 6 bidders			-1.13			-2.14*			-0.749
			(1.07)			(1.28)			(1.13)
Last month × 7+ bidders			-1.44			-2.27*			-1.44
			(1.04)			(1.21)			(1.11)
4 bidders		-0.236	-0.328		-2.52***	-2.53***		0.0257	-0.109
		(0.235)	(0.252)		(0.266)	(0.288)		(0.26)	(0.278)
5 bidders		-3.33***	-3.5***		-7.08***	-7.1***		-3.21***	-3.42***
		(0.307)	(0.316)		(0.358)	(0.383)		(0.328)	(0.34)
6 bidders		-6.06***	-5.91***		-10.5***	-10.2***		-5.97***	-5.88***
		(0.387)	(0.413)		(0.465)	(0.498)		(0.416)	(0.441)
7+ bidders		-11.1***	-11.0***		-13.1***	-12.9***		-11.9***	-11.7***
		(0.416)	(0.434)		(0.502)	(0.514)		(0.445)	(0.463)
R^2	0.291	0.298	0.298	0.406	0.416	0.416	0.323	0.331	0.331
N	519301	519301	519301	136751	136751	136751	382550	382550	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Columns 1, 4, and 7 correspond to specification 13. Columns 2, 5, and 8 also include dummies for each number of bidders in the auction. Columns 3, 6, and 9 further split the last month dummy by the number of bidders in the auction. Year FE stands for year fixed effects, buyer FE refers to government department fixed effects, seller FE indicates firm fixed effects, and sector FE represents sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

Table (J.2.6) The Year-End Effect for Margins (with Split by Year Interval)

	All 17-19	*All 20-21*	*Winners 17-19*	*Winners 20-21*	*Losers 17-19*	*Losers 20-21*
	(1)	(2)	(3)	(4)	(5)	(6)
Last month	2.24***	1.82***	2.08***	1.12**	2.52***	2.19***
	(0.512)	(0.518)	(0.609)	(0.541)	(0.594)	(0.591)
R^2	0.373	0.342	0.501	0.49	0.415	0.377
N	208764	310537	56309	80442	152455	230095
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation 12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

J.3 Specification 3

This is the benchmark specification but with all of the Leaky ReLU node activation functions used in each node being replaced instead by shrink hyperbolic tangent function ($x - \tanh(x)$).

Table (J.3.1) Final convergence

		R^2			Pearson Corr			MAE		
Seed	Num. Iterates	All	Test	Training	All	Test	Training	All	Test	Training
1	386	0.0632	0.0664	0.0629	0.2522	0.2587	0.2515	0.252	0.2516	0.252
2	1854	0.0753	0.0757	0.0752	0.2747	0.2759	0.2746	0.2495	0.2506	0.2494
3	200	0.0484	0.0475	0.0485	0.221	0.2193	0.2212	0.2549	0.2554	0.2548
4	321	0.0598	0.0589	0.0599	0.2455	0.2438	0.2457	0.2516	0.2518	0.2516
5	2395	0.0738	0.073	0.0739	0.2718	0.2704	0.272	0.251	0.251	0.251
6	950	0.0663	0.0652	0.0664	0.2584	0.2558	0.2587	0.2509	0.2502	0.251
7	1262	0.0734	0.0753	0.0732	0.2712	0.2746	0.2708	0.251	0.2501	0.251
8	539	0.0643	0.0631	0.0644	0.2542	0.2518	0.2545	0.2514	0.2525	0.2513
9	1744	0.0748	0.076	0.0747	0.2737	0.2758	0.2734	0.2511	0.2501	0.2512
10	298	0.0638	0.0617	0.064	0.253	0.2488	0.2535	0.2523	0.2517	0.2524

Note: This table shows the R^2, Pearson correlation and Mean Absolute Error (MAE) of the prediction at the point when training stops. Due to the early stopping logic, training stops after 50 iterates in which performance on the loss function on the test set has not improved.

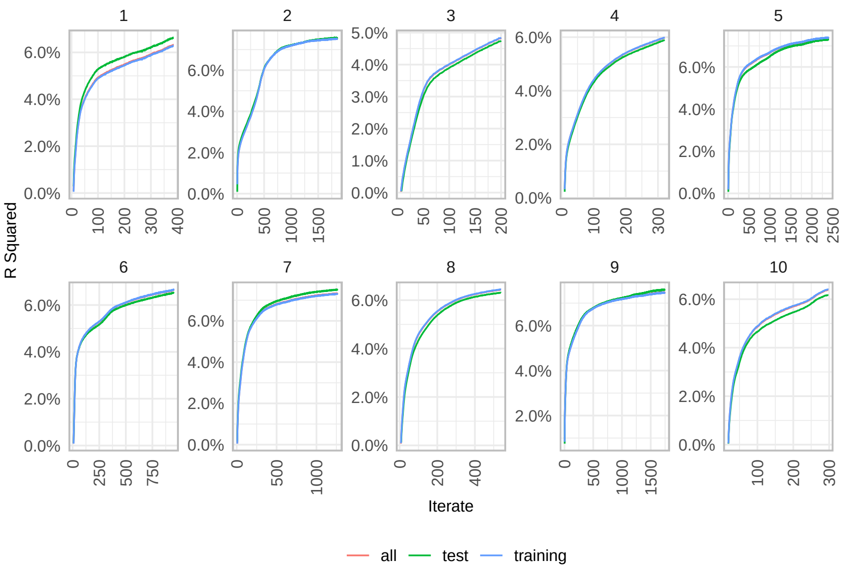
Table (J.3.2) Proportion of invalid values for each trained model.

Seed	Demand > 1	Demand < 0	Cost > Bid	Cost < 0
Combined	0.0021%	0.0021%	0%	0
1	0.0006%	0.0006%	0%	0
2	0.0021%	0.0021%	0.0004%	0
3	0.0005%	0.0005%	0%	0
4	0.0002%	0.0002%	0%	0
5	0.0022%	0.0022%	0%	0
6	0.0004%	0.0004%	0%	0
7	0.0034%	0.0034%	0.0246%	0
8	0.0005%	0.0005%	0.001%	0
9	0.0022%	0.0022%	0%	0
10	0.0009%	0.0009%	0%	0

Note: This table shows fraction of demands that are outside the unit interval and the fraction of costs that are above the final bid cost or below zero. Each figure in this table is a percentage of the non-missing values in the dataset. Figures are expressed to 4 decimal places.

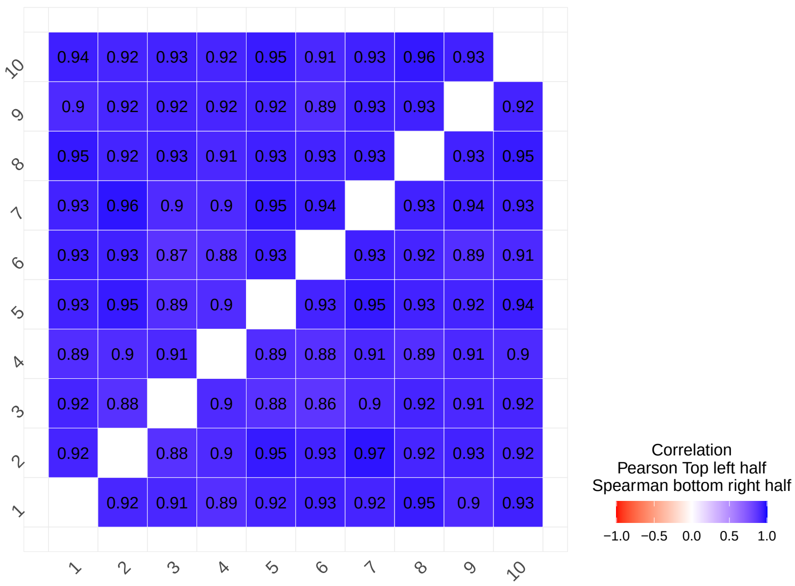
We first present tables and figures summarising the convergence path and fit of the model. The convergence plot of each seed with this specification can be seen in Figure J.3.1, while the correlation matrix between implied costs generated with each seed is displayed in Figure J.3.2. We summarise the final convergences achieved by each seed in Table J.3.1. The fraction of demands and costs that fall outside the feasible range is in Table J.3.2. Note that models that do not converge well may have more estimated probabilities outside the unit interval. Our analysis only uses the 5 trained models that have the highest R^2 however these poorly performing cases are not used. The variable importance of each of the model's input variables is summarised in Table J.3.3. We then present kernel densities of demand, cost, margin, o_1, o_2 and o_3 in Figure J.3.3. In Table J.3.4 (corresponding to Table 4) we provide the

Figure (J.3.1) Convergence Plot for Specification 3



Note: This chart plots the convergence of each trained model in terms of the R^2. We only plot positive R^2 for each which is why some seeds appear to start later on. In general, R^2 values start training at negative values.

Figure (J.3.2) Correlation Matrix Between Costs Implied by Each Seed for Specification 3



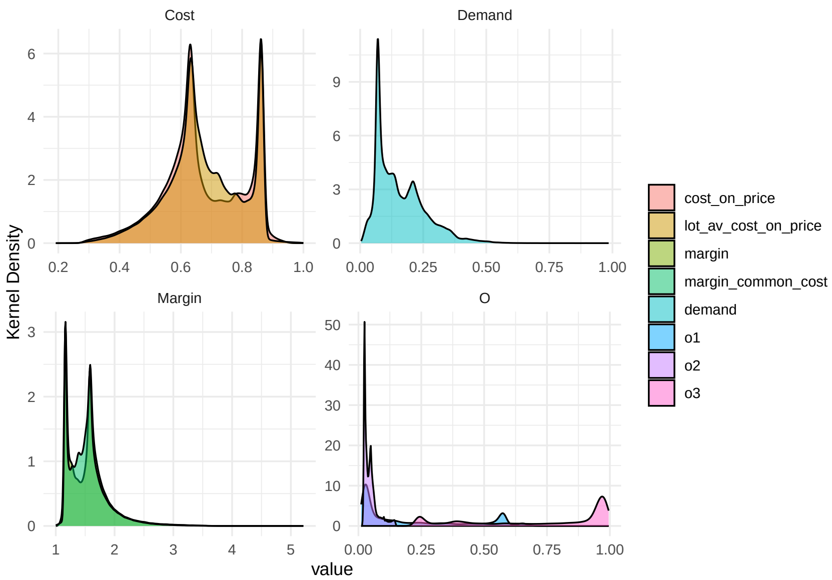
Note: This chart plots the correlation between $\frac{\text{Implied cost}}{\text{final price bid}}$ using the implied costs from each trained model. The top left half of the matrix are Pearson correlations while the bottom right half are Spearman correlations.

Table (J.3.3) Variable Importance Plot

Permuted Variable	Relative MSE	Relative Spearman Correlation
Price	9.9388	0.6563
Best Preceding Price	378.8362	0.7323
Log Expected Price	1.007	0.9603
Num. Bidders after	1.0623	0.5164
Mean prices bidder	1.0863	0.6182
Num. Participants	1.0467	0.7734
Prop. Inactive Bidders	1.0609	0.7643
Price Reduction in Round 3	1.0011	0.994
Round 2 std of bids	1.1577	0.4469
Year tender published	1.0031	0.9823

Note: To construct this table we took the 5 trained models (of 10 in total) with above median performance on the test set. We recorded it's performance on the full (test + training) sample as a benchmark. We then permuted each of the input variables (30 times per model) and recorded their performance. We recorded the relative MSE and Spearman correlation for each permutation and each model. We then took the medians for each model. Finally we took the median across models to get one relative performance metric per permuted variable. A relative MSE above 1 indicates a deterioration in performance by permuting that variable while a relative spearman below 1 indicates the same.

Figure (J.3.3) Kernel Densities of Values Estimated with Specification 3



Note: This chart plots the kernel densities of demand, implied margin, cost and the o_1, o_2, o_3 parameters. For the purposes of this figure, we remove observations where $\frac{\text{Implied cost}}{\text{final price bid}}$ is outside the $[0, 1]$ range or demand is outside this same range. The amount of observations that are dropped in this way are small and represented in Table J.3.2.

regression coefficients of margin with different fixed effects for winners and losers.

Table (J.3.4) The Year-End Effect for Margins (with Varying Fixed Effects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Winners								
Last month	2.68***	2.67***	2.15***	1.99***	2.26***	2.18***	2.06***	2.15***
	(0.447)	(0.446)	(0.44)	(0.406)	(0.44)	(0.412)	(0.441)	(0.44)
R^2	0.00157	0.0336	0.3	0.143	0.302	0.165	0.403	0.405
N	136756	136751	136756	136756	136751	136751	136756	136751
Panel B: Losers								
Last month	3.98***	3.81***	3.31***	3.26***	3.41***	3.31***	3.11***	3.21***
	(0.568)	(0.572)	(0.521)	(0.503)	(0.521)	(0.511)	(0.493)	(0.492)
R^2	0.00224	0.0398	0.232	0.129	0.235	0.155	0.322	0.325
N	382559	382550	382559	382559	382550	382550	382559	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	No	Yes	No	No	Yes	Yes	No	Yes
Seller FE	No	No	Yes	No	Yes	No	Yes	Yes
Buyer FE	No	No	No	Yes	No	Yes	Yes	Yes

Note: The sample is composed of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is a margin common cost as defined in equation 12. Last month is a dummy variable that takes the value one if the tender was finalised in December or November and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and by seller. *p<0.1; **p<0.05; ***p<0.01.

The regressions with the interaction terms are in Table J.3.5 (which corresponds to Table 5).

Finally in Table J.3.5 we replicate the regressions without controls on different time intervals of data. Specifically, for all bids, winners and losers respectively we perform separate regressions: one on 2017-2019 (pre-Covid period) and one on 2020-2021 (where the bidding process was impacted by covid). We present our findings in Table J.3.6.

Table (J.3.5) The Year-End Effect for Margins (with Varying Controls)

	All bids			*Winners*			*Losers*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Last month	2.75***	1.82***	1.84***	2.15***	1.33***	1.74***	3.21***	2.17***	2.04***
	(0.437)	(0.405)	(0.5)	(0.44)	(0.42)	(0.558)	(0.492)	(0.454)	(0.57)
Last month × 4 bidders			0.715			0.137			0.99
			(0.704)			(0.835)			(0.756)
Last month × 5 bidders			1.39			0.153			1.78*
			(0.945)			(1.16)			(0.995)
Last month × 6 bidders			-1.72			-2.68**			-1.36
			(1.14)			(1.36)			(1.21)
Last month × 7+ bidders			-2.42**			-3.35**			-2.41**
			(1.12)			(1.33)			(1.2)
4 bidders		-1.56***	-1.67***		-3.8***	-3.81***		-1.33***	-1.49***
		(0.259)	(0.277)		(0.292)	(0.317)		(0.286)	(0.306)
5 bidders		-7.39***	-7.59***		-10.8***	-10.8***		-7.38***	-7.64***
		(0.343)	(0.354)		(0.398)	(0.425)		(0.365)	(0.378)
6 bidders		-12.7***	-12.5***		-16.2***	-15.8***		-13.0***	-12.8***
		(0.434)	(0.461)		(0.521)	(0.555)		(0.463)	(0.489)
7+ bidders		-17.8***	-17.5***		-18.6***	-18.2***		-18.9***	-18.6***
		(0.468)	(0.481)		(0.569)	(0.582)		(0.495)	(0.507)
R^2	0.294	0.31	0.31	0.405	0.422	0.423	0.325	0.342	0.342
N	519301	519301	519301	136751	136751	136751	382550	382550	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Columns 1, 4, and 7 correspond to specification 13. Columns 2, 5, and 8 also include dummies for each number of bidders in the auction. Columns 3, 6, and 9 further split the last month dummy by the number of bidders in the auction. Year FE stands for year fixed effects, buyer FE refers to government department fixed effects, seller FE indicates firm fixed effects, and sector FE represents sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

Table (J.3.6) The Year-End Effect for Margins (with Split by Year Interval)

	All 17-19	*All 20-21*	*Winners 17-19*	*Winners 20-21*	*Losers 17-19*	*Losers 20-21*
	(1)	(2)	(3)	(4)	(5)	(6)
Last month	2.27***	2.69***	2.16***	1.87***	2.5***	3.09***
	(0.561)	(0.576)	(0.676)	(0.596)	(0.644)	(0.653)
R^2	0.378	0.345	0.502	0.489	0.418	0.378
N	208764	310537	56309	80442	152455	230095
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation 12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

J.4 Specification 4

This is the benchmark specification but with all of the Leaky ReLU node activation functions used in each node being replaced instead by swish (Ramachandran, Zoph and Le, 2017) functions.

Table (J.4.1) Final convergence

		R^2			Pearson Corr			MAE		
Seed	Num. Iterates	All	Test	Training	All	Test	Training	All	Test	Training
1	98	0.0377	0.04	0.0374	0.1951	0.2012	0.1944	0.2592	0.2591	0.2592
2	502	0.0632	0.0618	0.0633	0.2521	0.2497	0.2523	0.252	0.2534	0.2519
3	193	0.0533	0.0529	0.0533	0.2314	0.2309	0.2315	0.255	0.2555	0.255
4	565	0.0642	0.0625	0.0644	0.2552	0.2517	0.2555	0.2507	0.251	0.2506
5	452	0.0574	0.0561	0.0576	0.2431	0.2408	0.2433	0.2633	0.2634	0.2633
6	1402	0.0714	0.07	0.0715	0.2678	0.2649	0.2681	0.2515	0.2509	0.2516
7	624	0.0677	0.068	0.0676	0.2615	0.2621	0.2614	0.2504	0.2498	0.2504
8	751	0.0662	0.0654	0.0663	0.2576	0.2562	0.2578	0.2526	0.2536	0.2524
9	382	0.0642	0.0651	0.0641	0.2538	0.2555	0.2536	0.2533	0.2524	0.2534
10	1538	0.0687	0.0657	0.069	0.2627	0.2566	0.2634	0.251	0.2505	0.2511

Note: This table shows the R^2, Pearson correlation and Mean Absolute Error (MAE) of the prediction at the point when training stops. Due to the early stopping logic, training stops after 50 iterates in which performance on the loss function on the test set has not improved.

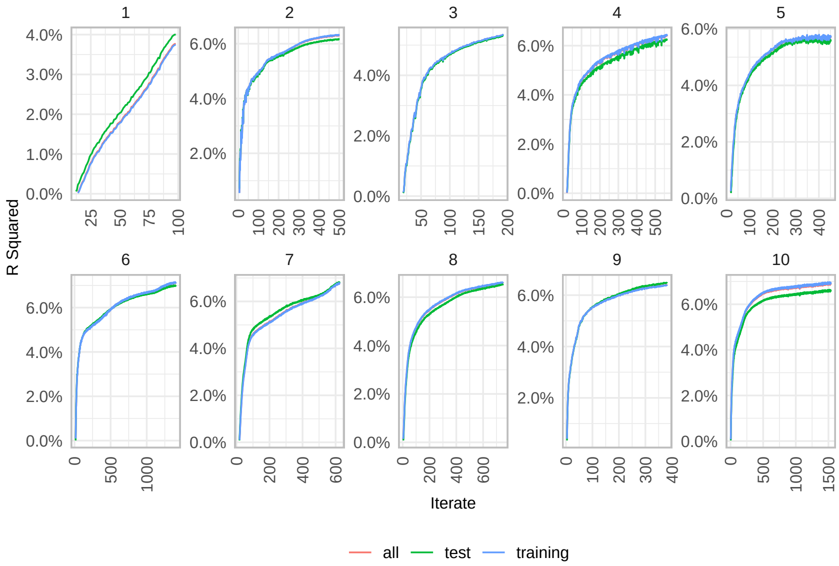
Table (J.4.2) Proportion of invalid values for each trained model.

Seed	Demand > 1	Demand < 0	Cost > Bid	Cost < 0
Combined	0.0009%	0.0009%	0.0708%	0
1	0.0005%	0.0005%	0%	0
2	0.0001%	0.0001%	0.0776%	0
3	0.0004%	0.0004%	0%	0
4	0.0017%	0.0017%	0.292%	0
5	0.0004%	0.0004%	0.0087%	0
6	0.0006%	0.0006%	0.0565%	0
7	0.0018%	0.0018%	0.1826%	0
8	0.0002%	0.0002%	0.3395%	0
9	0.0002%	0.0002%	0.3267%	0
10	0.0037%	0.0037%	0.0484%	0

Note: This table shows fraction of demands that are outside the unit interval and the fraction of costs that are above the final bid cost or below zero. Each figure in this table is a percentage of the non-missing values in the dataset. Figures are expressed to 4 decimal places.

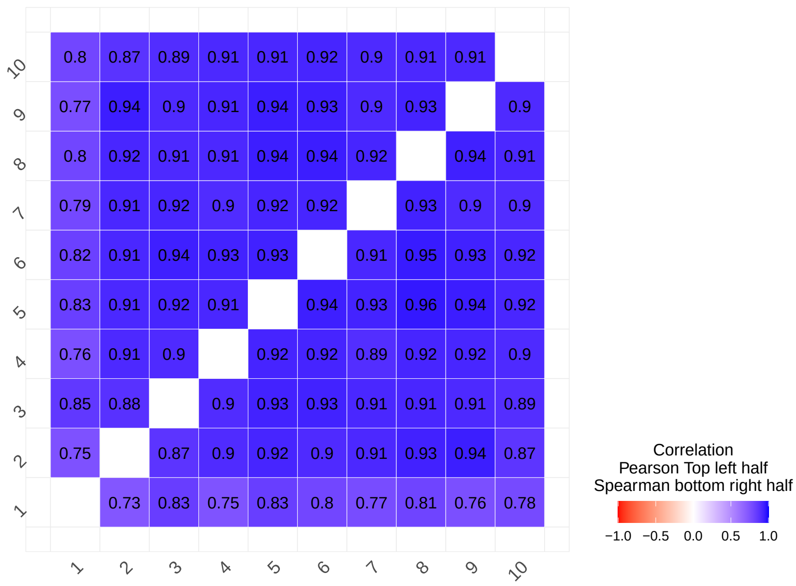
We first present tables and figures summarising the convergence path and fit of the model. The convergence plot of each seed with this specification can be seen in Figure J.4.1, while the correlation matrix between implied costs generated with each seed is displayed in Figure J.4.2. We summarise the final convergences achieved by each seed in Table J.4.1. The fraction of demands and costs that fall outside the feasible range is in Table J.4.2. Note that models that do not converge well may have more estimated probabilities outside the unit interval. Our analysis only uses the 5 trained models that have the highest R^2 however these poorly performing cases are not used. The variable importance of each of the model's input variables is summarised in Table J.4.3. We then present kernel densities of demand, cost, margin, o_1, o_2 and o_3 in Figure J.4.3. In Table J.4.4 (corresponding to Table 4) we provide the

Figure (J.4.1) Convergence Plot for Specification 4



Note: This chart plots the convergence of each trained model in terms of the R^2. We only plot positive R^2 for each which is why some seeds appear to start later on. In general, R^2 values start training at negative values.

Figure (J.4.2) Correlation Matrix Between Costs Implied by Each Seed for Specification 4



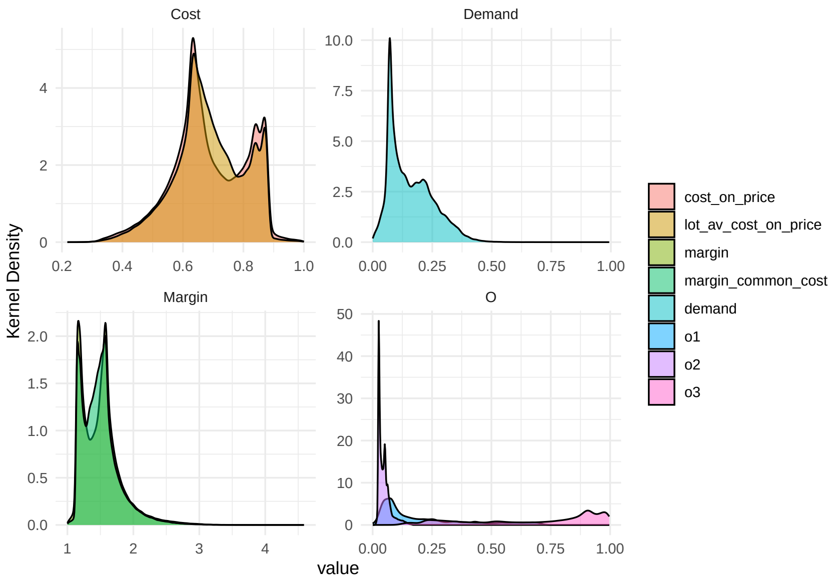
Note: This chart plots the correlation between $\frac{\text{Implied cost}}{\text{final price bid}}$ using the implied costs from each trained model. The top left half of the matrix are Pearson correlations while the bottom right half are Spearman correlations.

Table (J.4.3) Variable Importance Plot

Permuted Variable	Relative MSE	Relative Spearman Correlation
Price	12.1884	0.6408
Best Preceding Price	140.755	0.704
Log Expected Price	1.0077	0.9496
Num. Bidders after	1.0618	0.4941
Mean prices bidder	1.0675	0.591
Num. Participants	1.0454	0.7171
Prop. Inactive Bidders	1.0363	0.7633
Price Reduction in Round 3	1.0011	0.9935
Round 2 std of bids	1.1461	0.4389
Year tender published	1.003	0.9812

Note: To construct this table we took the 5 trained models (of 10 in total) with above median performance on the test set. We recorded it's performance on the full (test + training) sample as a benchmark. We then permuted each of the input variables (30 times per model) and recorded their performance. We recorded the relative MSE and Spearman correlation for each permutation and each model. We then took the medians for each model. Finally we took the median across models to get one relative performance metric per permuted variable. A relative MSE above 1 indicates a deterioration in performance by permuting that variable while a relative spearman below 1 indicates the same.

Figure (J.4.3) Kernel Densities of Values Estimated with Specification 4



Note: This chart plots the kernel densities of demand, implied margin, cost and the o_1, o_2, o_3 parameters. For the purposes of this figure, we remove observations where $\frac{\text{Implied cost}}{\text{final price bid}}$ is outside the $[0, 1]$ range or demand is outside this same range. The amount of observations that are dropped in this way are small and represented in Table J.4.2.

regression coefficients of margin with different fixed effects for winners and losers.

Table (J.4.4) The Year-End Effect for Margins (with Varying Fixed Effects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Winners								
Last month	2.54***	2.43***	1.97***	1.91***	2.06***	2.0***	1.88***	1.96***
	(0.393)	(0.387)	(0.386)	(0.36)	(0.385)	(0.362)	(0.387)	(0.387)
R^2	0.00123	0.0302	0.298	0.143	0.3	0.163	0.402	0.404
N	136756	136751	136756	136756	136751	136751	136756	136751
Panel B: Losers								
Last month	3.93***	3.59***	3.15***	3.26***	3.23***	3.17***	3.01***	3.09***
	(0.522)	(0.51)	(0.465)	(0.454)	(0.465)	(0.455)	(0.442)	(0.442)
R^2	0.00178	0.0338	0.226	0.125	0.228	0.146	0.315	0.317
N	382559	382550	382559	382559	382550	382550	382559	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	No	Yes	No	No	Yes	Yes	No	Yes
Seller FE	No	No	Yes	No	Yes	No	Yes	Yes
Buyer FE	No	No	No	Yes	No	Yes	Yes	Yes

Note: The sample is composed of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is a margin common cost as defined in equation 12. Last month is a dummy variable that takes the value one if the tender was finalised in December or November and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and by seller. *p<0.1; **p<0.05; ***p<0.01.

The regressions with the interaction terms are in Table J.4.5 (which corresponds to Table 5).

Finally in Table J.4.5 we replicate the regressions without controls on different time intervals of data. Specifically, for all bids, winners and losers respectively we perform separate regressions: one on 2017-2019 (pre-Covid period) and one on 2020-2021 (where the bidding process was impacted by covid). We present our findings in Table J.4.6.

Table (J.4.5) The Year-End Effect for Margins (with Varying Controls)

	All bids			*Winners*			*Losers*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Last month	2.63***	1.75***	1.63***	1.96***	1.19***	1.49***	3.09***	2.1***	1.84***
	(0.389)	(0.36)	(0.45)	(0.387)	(0.369)	(0.498)	(0.442)	(0.405)	(0.513)
Last month × 4 bidders			0.512			0.0411			0.714
			(0.624)			(0.742)			(0.671)
Last month × 5 bidders			1.1			0.154			1.39
			(0.822)			(1.01)			(0.869)
Last month × 6 bidders			-1.01			-2.41**			-0.603
			(1.03)			(1.19)			(1.09)
Last month × 7+ bidders			-0.919			-1.97*			-0.854
			(0.994)			(1.15)			(1.07)
4 bidders		-1.75***	-1.83***		-3.95***	-3.96***		-1.49***	-1.6***
		(0.233)	(0.249)		(0.258)	(0.281)		(0.257)	(0.275)
5 bidders		-5.99***	-6.15***		-9.41***	-9.43***		-5.92***	-6.12***
		(0.299)	(0.309)		(0.345)	(0.37)		(0.319)	(0.332)
6 bidders		-10.4***	-10.3***		-13.9***	-13.6***		-10.5***	-10.4***
		(0.385)	(0.409)		(0.455)	(0.486)		(0.413)	(0.437)
7+ bidders		-17.8***	-17.7***		-18.4***	-18.2***		-18.7***	-18.7***
		(0.425)	(0.441)		(0.496)	(0.512)		(0.455)	(0.472)
R^2	0.286	0.305	0.305	0.404	0.423	0.423	0.317	0.337	0.337
N	519301	519301	519301	136751	136751	136751	382550	382550	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Columns 1, 4, and 7 correspond to specification 13. Columns 2, 5, and 8 also include dummies for each number of bidders in the auction. Columns 3, 6, and 9 further split the last month dummy by the number of bidders in the auction. Year FE stands for year fixed effects, buyer FE refers to government department fixed effects, seller FE indicates firm fixed effects, and sector FE represents sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

Table (J.4.6) The Year-End Effect for Margins (with Split by Year Interval)

	All 17-19	*All 20-21*	*Winners 17-19*	*Winners 20-21*	*Losers 17-19*	*Losers 20-21*
	(1)	(2)	(3)	(4)	(5)	(6)
Last month	2.29***	2.59***	1.94***	1.77***	2.6***	2.99***
	(0.498)	(0.515)	(0.6)	(0.528)	(0.576)	(0.586)
R^2	0.373	0.336	0.501	0.488	0.414	0.369
N	208764	310537	56309	80442	152455	230095
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation 12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

J.5 Specification 5

This specification is the same as the benchmark specification but with alternative layers of different neurons in the neural network. The input layer and 3rd and 6th hidden layers are leaky ReLU, the 1st and 4th hidden layers are $x - \tanh(x)$ and the 2nd and 5th hidden layers are swish (Ramachandran, Zoph and Le, 2017).

Table (J.5.1) Final convergence

		R^2			Pearson Corr			MAE		
Seed	Num. Iterates	All	Test	Training	All	Test	Training	All	Test	Training
1	134	0.0452	0.047	0.045	0.2128	0.2172	0.2123	0.259	0.2589	0.259
2	307	0.0566	0.056	0.0567	0.2385	0.2376	0.2386	0.2554	0.2567	0.2553
3	865	0.066	0.065	0.0661	0.2577	0.256	0.2578	0.2516	0.2521	0.2515
4	466	0.0632	0.0616	0.0634	0.252	0.2486	0.2524	0.2552	0.2555	0.2552
5	984	0.0681	0.0668	0.0682	0.2613	0.2588	0.2616	0.2572	0.2573	0.2572
6	720	0.0641	0.063	0.0642	0.2555	0.2528	0.2559	0.2519	0.2512	0.2519
7	1326	0.0685	0.0692	0.0685	0.2641	0.2655	0.264	0.2505	0.2499	0.2506
8	1394	0.0658	0.0656	0.0658	0.2595	0.2592	0.2595	0.2504	0.2513	0.2502
9	159	0.0543	0.0552	0.0542	0.2349	0.2363	0.2348	0.2524	0.2516	0.2525
10	1517	0.0735	0.0702	0.0739	0.2715	0.2651	0.2722	0.2517	0.2512	0.2517

Note: This table shows the R^2, Pearson correlation and Mean Absolute Error (MAE) of the prediction at the point when training stops. Due to the early stopping logic, training stops after 50 iterates in which performance on the loss function on the test set has not improved.

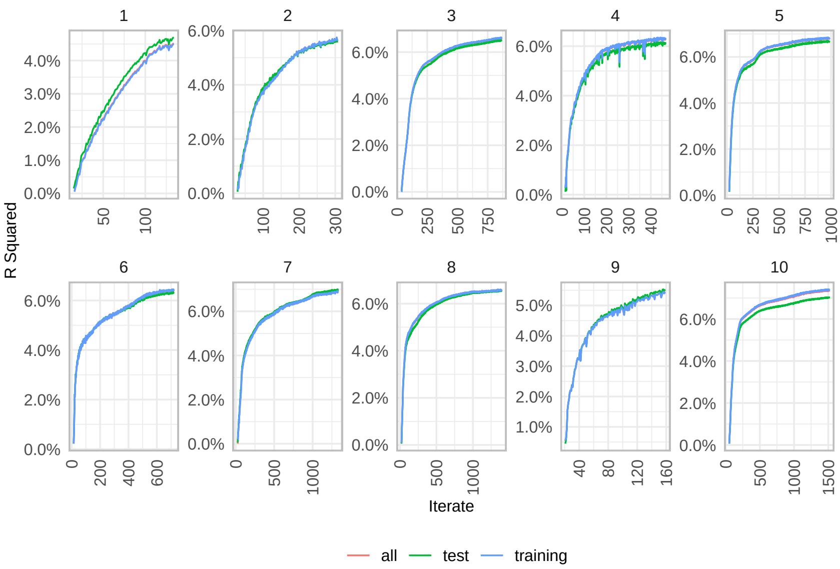
Table (J.5.2) Proportion of invalid values for each trained model.

Seed	Demand > 1	Demand < 0	Cost > Bid	Cost < 0
Combined	0.001%	0.001%	0.0142%	0
1	0.0001%	0.0001%	0%	0
2	0%	0%	0%	0
3	0.0002%	0.0002%	0.1083%	0
4	0%	0%	0.1072%	0
5	0.0007%	0.0007%	0%	0
6	0.0004%	0.0004%	0.1353%	0
7	0.0005%	0.0005%	0.1078%	0
8	0.0006%	0.0006%	0.0147%	0
9	0.0001%	0.0001%	0%	0
10	0.0011%	0.0011%	0%	0

Note: This table shows fraction of demands that are outside the unit interval and the fraction of costs that are above the final bid cost or below zero. Each figure in this table is a percentage of the non-missing values in the dataset. Figures are expressed to 4 decimal places.

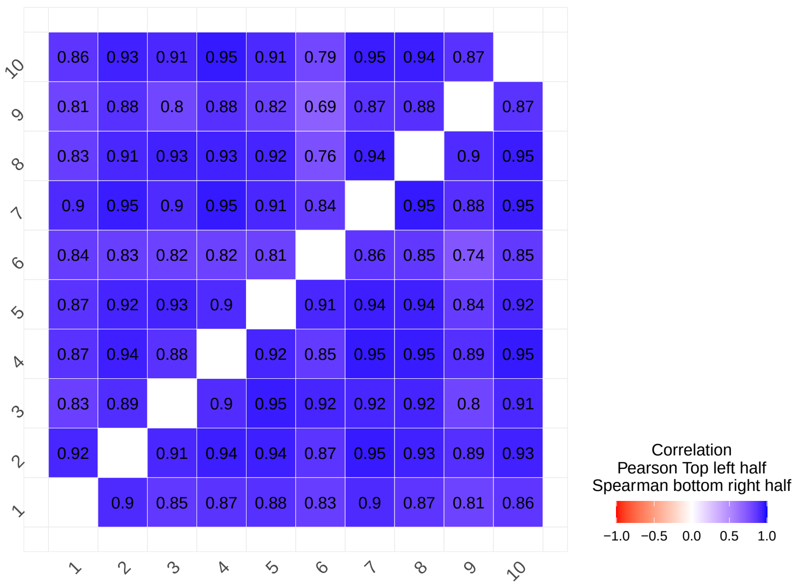
We first present tables and figures summarising the convergence path and fit of the model. The convergence plot of each seed with this specification can be seen in Figure J.5.1, while the correlation matrix between implied costs generated with each seed is displayed in Figure J.5.2. We summarise the final convergences achieved by each seed in Table J.5.1. The fraction of demands and costs that fall outside the feasible range is in Table J.5.2. Note that models that do not converge well may have more estimated probabilities outside the unit interval. Our analysis only uses the 5 trained models that have

Figure (J.5.1) Convergence Plot for Specification 5



Note: This chart plots the convergence of each trained model in terms of the R^2. We only plot positive R^2 for each which is why some seeds appear to start later on. In general, R^2 values start training at negative values.

Figure (J.5.2) Correlation Matrix Between Costs Implied by Each Seed for Specification 5



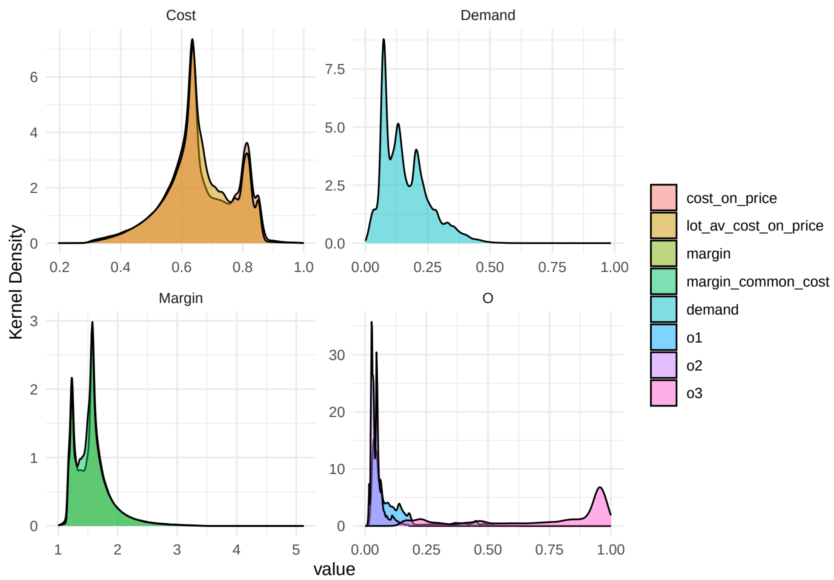
Note: This chart plots the correlation between $\frac{\text{Implied cost}}{\text{final price bid}}$ using the implied costs from each trained model. The top left half of the matrix are Pearson correlations while the bottom right half are Spearman correlations.

Table (J.5.3) Variable Importance Plot

Permuted Variable	Relative MSE	Relative Spearman Correlation
Price	12.3247	0.6694
Best Preceding Price	8.5039	0.7401
Log Expected Price	1.0069	0.9536
Num. Bidders after	1.0556	0.5312
Mean prices bidder	1.0739	0.5658
Num. Participants	1.0387	0.8067
Prop. Inactive Bidders	1.0513	0.689
Price Reduction in Round 3	1.0018	0.9885
Round 2 std of bids	1.1436	0.463
Year tender published	1.0028	0.9829

Note: To construct this table we took the 5 trained models (of 10 in total) with above median performance on the test set. We recorded it's performance on the full (test + training) sample as a benchmark. We then permuted each of the input variables (30 times per model) and recorded their performance. We recorded the relative MSE and Spearman correlation for each permutation and each model. We then took the medians for each model. Finally we took the median across models to get one relative performance metric per permuted variable. A relative MSE above 1 indicates a deterioration in performance by permuting that variable while a relative spearman below 1 indicates the same.

Figure (J.5.3) Kernel Densities of Values Estimated with Specification 5



Note: This chart plots the kernel densities of demand, implied margin, cost and the o_1, o_2, o_3 parameters. For the purposes of this figure, we remove observations where $\frac{\text{Implied cost}}{\text{final price bid}}$ is outside the $[0, 1]$ range or demand is outside this same range. The amount of observations that are dropped in this way are small and represented in Table J.5.2.

the highest R^2 however these poorly performing cases are not used. The variable importance of each of the model's input variables is summarised in Table J.5.3. We then present kernel densities of demand, cost, margin, o_1, o_2 and o_3 in Figure J.5.3. In Table J.5.4 (corresponding to Table 4) we provide the regression coefficients of margin with different fixed effects for winners and losers.

Table (J.5.4) The Year-End Effect for Margins (with Varying Fixed Effects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Winners								
Last month	2.42***	2.28***	1.83***	1.75***	1.92***	1.83***	1.71***	1.79***
	(0.432)	(0.419)	(0.408)	(0.382)	(0.407)	(0.38)	(0.401)	(0.402)
R^2	0.00144	0.0317	0.302	0.142	0.305	0.164	0.406	0.408
N	136756	136751	136756	136756	136751	136751	136756	136751
Panel B: Losers								
Last month	3.84***	3.47***	3.02***	3.12***	3.11***	3.02***	2.81***	2.9***
	(0.573)	(0.552)	(0.503)	(0.488)	(0.503)	(0.484)	(0.471)	(0.47)
R^2	0.00265	0.0373	0.233	0.126	0.236	0.151	0.321	0.324
N	382559	382550	382559	382559	382550	382550	382559	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	No	Yes	No	No	Yes	Yes	No	Yes
Seller FE	No	No	Yes	No	Yes	No	Yes	Yes
Buyer FE	No	No	No	Yes	No	Yes	Yes	Yes

Note: The sample is composed of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is a margin common cost as defined in equation 12. Last month is a dummy variable that takes the value one if the tender was finalised in December or November and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and by seller. *p<0.1; **p<0.05; ***p<0.01.

The regressions with the interaction terms are in Table J.5.5 (which corresponds to Table 5).

Finally in Table J.5.5 we replicate the regressions without controls on different time intervals of data. Specifically, for all bids, winners and losers respectively we perform separate regressions: one on 2017-2019 (pre-Covid period) and one on 2020-2021 (where the bidding process was impacted by covid). We present our findings in Table J.5.6.

Table (J.5.5) The Year-End Effect for Margins (with Varying Controls)

	All bids			*Winners*			*Losers*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Last month	2.43***	1.63***	1.5***	1.79***	1.08***	1.37***	2.9***	2.01***	1.7***
	(0.411)	(0.383)	(0.468)	(0.402)	(0.385)	(0.517)	(0.47)	(0.435)	(0.536)
Last month × 4 bidders			0.468			-0.0666			0.727
			(0.644)			(0.77)			(0.693)
Last month × 5 bidders			1.03			0.0811			1.35
			(0.858)			(1.08)			(0.907)
Last month × 6 bidders			-0.528			-1.76			-0.0998
			(1.06)			(1.27)			(1.13)
Last month × 7+ bidders			-0.957			-1.91			-0.877
			(1.01)			(1.2)			(1.09)
4 bidders		-0.249	-0.323		-2.64***	-2.62***		0.0501	-0.0662
		(0.237)	(0.255)		(0.267)	(0.29)		(0.262)	(0.282)
5 bidders		-2.85***	-3.0***		-6.78***	-6.79***		-2.68***	-2.88***
		(0.307)	(0.317)		(0.356)	(0.381)		(0.328)	(0.342)
6 bidders		-7.16***	-7.1***		-11.5***	-11.2***		-7.09***	-7.09***
		(0.393)	(0.418)		(0.466)	(0.498)		(0.424)	(0.447)
7+ bidders		-16.5***	-16.4***		-17.6***	-17.4***		-17.4***	-17.3***
		(0.434)	(0.451)		(0.514)	(0.527)		(0.466)	(0.485)
R^2	0.292	0.307	0.307	0.408	0.423	0.423	0.324	0.34	0.34
N	519301	519301	519301	136751	136751	136751	382550	382550	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Columns 1, 4, and 7 correspond to specification 13. Columns 2, 5, and 8 also include dummies for each number of bidders in the auction. Columns 3, 6, and 9 further split the last month dummy by the number of bidders in the auction. Year FE stands for year fixed effects, buyer FE refers to government department fixed effects, seller FE indicates firm fixed effects, and sector FE represents sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

Table (J.5.6) The Year-End Effect for Margins (with Split by Year Interval)

	All 17-19	*All 20-21*	*Winners 17-19*	*Winners 20-21*	*Losers 17-19*	*Losers 20-21*
	(1)	(2)	(3)	(4)	(5)	(6)
Last month	2.39***	2.1***	2.06***	1.29**	2.73***	2.52***
	(0.525)	(0.532)	(0.623)	(0.546)	(0.605)	(0.608)
R^2	0.374	0.344	0.501	0.493	0.415	0.378
N	208764	310537	56309	80442	152455	230095
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation 12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

J.6 Specification 6

This is the benchmark specification but with a cost floor of 0.0. Note that to avoid division by numbers close to zero we use only inferred costs that are at least 20% of the sales price (so all margins used in regressions are in the [0,5] range).

Table (J.6.1) Final convergence

		R^2			Pearson Corr			MAE		
Seed	Num. Iterates	All	Test	Training	All	Test	Training	All	Test	Training
1	54	0.0281	0.0299	0.0279	0.1695	0.1749	0.1689	0.26	0.26	0.26
2	176	0.0542	0.0535	0.0543	0.2348	0.2337	0.2349	0.2552	0.2564	0.2551
3	814	0.0689	0.0683	0.069	0.2631	0.262	0.2632	0.253	0.2534	0.2529
4	127	0.0413	0.0407	0.0413	0.2038	0.2025	0.204	0.2586	0.2588	0.2586
5	272	0.0588	0.0582	0.0589	0.2431	0.2419	0.2432	0.2614	0.2614	0.2614
6	721	0.066	0.0646	0.0662	0.2571	0.2546	0.2574	0.2577	0.2571	0.2578
7	517	0.0591	0.0615	0.0588	0.251	0.2569	0.2504	0.249	0.2482	0.2491
8	349	0.0623	0.0605	0.0624	0.2534	0.2497	0.2538	0.2518	0.2528	0.2516
9	153	0.0564	0.0571	0.0563	0.2386	0.2399	0.2384	0.2539	0.2531	0.254
10	343	0.0609	0.0591	0.0611	0.2485	0.2442	0.249	0.2547	0.254	0.2548

Note: This table shows the R^2, Pearson correlation and Mean Absolute Error (MAE) of the prediction at the point when training stops. Due to the early stopping logic, training stops after 50 iterates in which performance on the loss function on the test set has not improved.

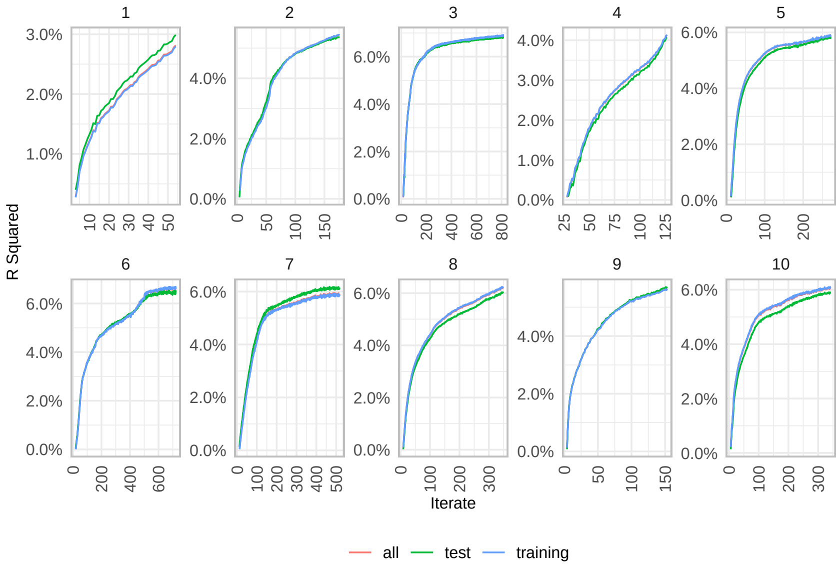
Table (J.6.2) Proportion of invalid values for each trained model.

Seed	Demand > 1	Demand < 0	Cost > Bid	Cost < 0
Combined	0%	0%	0.0012%	0
1	0.0006%	0.0006%	0%	0
2	0%	0%	0%	0
3	0%	0%	0.157%	0
4	0%	0%	0%	0
5	0%	0%	0.0002%	0
6	0%	0%	0.1071%	0
7	0%	0%	0.0012%	0
8	0.0006%	0.0006%	0%	0
9	0%	0%	0%	0
10	0%	0%	0%	0

Note: This table shows fraction of demands that are outside the unit interval and the fraction of costs that are above the final bid cost or below zero. Each figure in this table is a percentage of the non-missing values in the dataset. Figures are expressed to 4 decimal places.

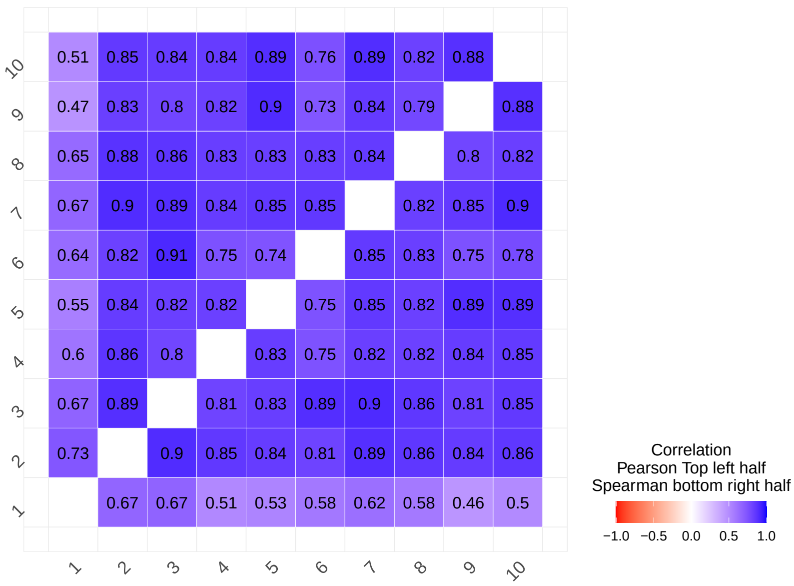
We first present tables and figures summarising the convergence path and fit of the model. The convergence plot of each seed with this specification can be seen in Figure J.6.1, while the correlation matrix between implied costs generated with each seed is displayed in Figure J.6.2. We summarise the final convergences achieved by each seed in Table J.6.1. The fraction of demands and costs that fall outside the feasible range is in Table J.6.2. Note that models that do not converge well may have more estimated probabilities outside the unit interval. Our analysis only uses the 5 trained models that have the highest R^2 however these poorly performing cases are not used. The variable importance of each of

Figure (J.6.1) Convergence Plot for Specification 6



Note: This chart plots the convergence of each trained model in terms of the R^2. We only plot positive R^2 for each which is why some seeds appear to start later on. In general, R^2 values start training at negative values.

Figure (J.6.2) Correlation Matrix Between Costs Implied by Each Seed for Specification 6



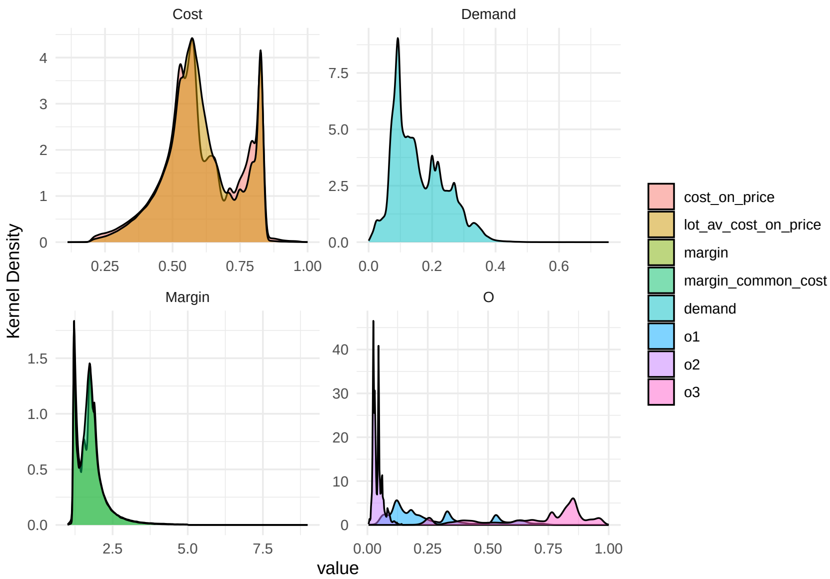
Note: This chart plots the correlation between $\frac{\text{Implied cost}}{\text{final price bid}}$ using the implied costs from each trained model. The top left half of the matrix are Pearson correlations while the bottom right half are Spearman correlations.

Table (J.6.3) Variable Importance Plot

Permuted Variable	Relative MSE	Relative Spearman Correlation
Price	1.7878	0.6775
Best Preceding Price	20.2356	0.7415
Log Expected Price	1.0057	0.9593
Num. Bidders after	1.0557	0.5067
Mean prices bidder	1.0539	0.5632
Num. Participants	1.0262	0.8004
Prop. Inactive Bidders	1.0457	0.6892
Price Reduction in Round 3	1.0014	0.9909
Round 2 std of bids	1.1098	0.4513
Year tender published	1.0028	0.9799

Note: To construct this table we took the 5 trained models (of 10 in total) with above median performance on the test set. We recorded it's performance on the full (test + training) sample as a benchmark. We then permuted each of the input variables (30 times per model) and recorded their performance. We recorded the relative MSE and Spearman correlation for each permutation and each model. We then took the medians for each model. Finally we took the median across models to get one relative performance metric per permuted variable. A relative MSE above 1 indicates a deterioration in performance by permuting that variable while a relative spearman below 1 indicates the same.

Figure (J.6.3) Kernel Densities of Values Estimated with Specification 6



Note: This chart plots the kernel densities of demand, implied margin, cost and the o_1, o_2, o_3 parameters. For the purposes of this figure, we remove observations where $\frac{\text{Implied cost}}{\text{final price bid}}$ is outside the $[0, 1]$ range or demand is outside this same range. The amount of observations that are dropped in this way are small and represented in Table J.6.2.

the model's input variables is summarised in Table J.6.3. We then present kernel densities of demand, cost, margin, o_1, o_2 and o_3 in Figure J.6.3. In Table J.6.4 (corresponding to Table 4) we provide the regression coefficients of margin with different fixed effects for winners and losers.

Table (J.6.4) The Year-End Effect for Margins (with Varying Fixed Effects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Winners								
Last month	4.54***	4.5***	3.46***	3.45***	3.61***	3.68***	3.28***	3.42***
	(0.773)	(0.75)	(0.722)	(0.681)	(0.721)	(0.68)	(0.719)	(0.718)
R^2	0.0019	0.0321	0.299	0.14	0.302	0.162	0.4	0.403
N	136756	136751	136756	136756	136751	136751	136756	136751
Panel B: Losers								
Last month	6.08***	5.79***	5.01***	5.04***	5.17***	5.06***	4.69***	4.85***
	(0.993)	(0.96)	(0.873)	(0.838)	(0.873)	(0.831)	(0.81)	(0.808)
R^2	0.00226	0.0386	0.233	0.128	0.237	0.154	0.322	0.325
N	382559	382550	382559	382559	382550	382550	382559	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	No	Yes	No	No	Yes	Yes	No	Yes
Seller FE	No	No	Yes	No	Yes	No	Yes	Yes
Buyer FE	No	No	No	Yes	No	Yes	Yes	Yes

Note: The sample is composed of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is a margin common cost as defined in equation 12. Last month is a dummy variable that takes the value one if the tender was finalised in December or November and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and by seller. *p<0.1; **p<0.05; ***p<0.01.

The regressions with the interaction terms are in Table J.6.5 (which corresponds to Table 5).

Finally in Table J.6.5 we replicate the regressions without controls on different time intervals of data. Specifically, for all bids, winners and losers respectively we perform separate regressions: one on 2017-2019 (pre-Covid period) and one on 2020-2021 (where the bidding process was impacted by covid). We present our findings in Table J.6.6.

Table (J.6.5) The Year-End Effect for Margins (with Varying Controls)

	All bids			*Winners*			*Losers*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Last month	4.18***	3.12***	3.4***	3.42***	2.43***	3.23***	4.85***	3.67***	3.7***
	(0.714)	(0.676)	(0.877)	(0.718)	(0.694)	(0.948)	(0.808)	(0.761)	(1.0)
Last month × 4 bidders			0.579			-0.285			1.07
			(1.14)			(1.36)			(1.22)
Last month × 5 bidders			1.44			0.188			1.99
			(1.5)			(1.89)			(1.57)
Last month × 6 bidders			-3.14*			-4.7**			-2.55
			(1.68)			(2.05)			(1.79)
Last month × 7+ bidders			-3.32**			-5.01**			-3.18*
			(1.63)			(1.96)			(1.75)
4 bidders		-0.0381	-0.127		-3.42***	-3.37***		0.43	0.262
		(0.404)	(0.433)		(0.463)	(0.501)		(0.439)	(0.469)
5 bidders		-4.36***	-4.56***		-9.42***	-9.44***		-4.12***	-4.41***
		(0.505)	(0.52)		(0.607)	(0.647)		(0.533)	(0.551)
6 bidders		-10.5***	-10.1***		-16.7***	-16.0***		-10.2***	-9.87***
		(0.629)	(0.668)		(0.746)	(0.803)		(0.671)	(0.707)
7+ bidders		-21.4***	-21.0***		-23.8***	-23.2***		-22.4***	-22.0***
		(0.644)	(0.658)		(0.778)	(0.799)		(0.692)	(0.702)
R^2	0.295	0.304	0.304	0.403	0.412	0.412	0.325	0.335	0.335
N	519301	519301	519301	136751	136751	136751	382550	382550	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Columns 1, 4, and 7 correspond to specification 13. Columns 2, 5, and 8 also include dummies for each number of bidders in the auction. Columns 3, 6, and 9 further split the last month dummy by the number of bidders in the auction. Year FE stands for year fixed effects, buyer FE refers to government department fixed effects, seller FE indicates firm fixed effects, and sector FE represents sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

Table (J.6.6) The Year-End Effect for Margins (with Split by Year Interval)

	All 17-19	*All 20-21*	*Winners 17-19*	*Winners 20-21*	*Losers 17-19*	*Losers 20-21*
	(1)	(2)	(3)	(4)	(5)	(6)
Last month	4.77***	3.16***	4.51***	2.13**	5.14***	3.74***
	(0.916)	(0.882)	(1.12)	(0.928)	(1.04)	(0.996)
R^2	0.383	0.347	0.5	0.488	0.422	0.379
N	208764	310537	56309	80442	152455	230095
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation 12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

J.7 Specification 7

This is the benchmark specification but with a cost floor of 0.2 rather than 0.25.

Table (J.7.1) Final convergence

		R^2			Pearson Corr			MAE		
Seed	Num. Iterates	All	Test	Training	All	Test	Training	All	Test	Training
1	89	0.0359	0.038	0.0357	0.1919	0.1976	0.1913	0.2572	0.2571	0.2572
2	195	0.0548	0.0542	0.0549	0.236	0.2353	0.2361	0.2534	0.2546	0.2532
3	323	0.0619	0.0615	0.062	0.2496	0.249	0.2497	0.2527	0.2532	0.2527
4	817	0.0623	0.0606	0.0624	0.2503	0.2468	0.2507	0.2581	0.2584	0.2581
5	112	0.0472	0.0457	0.0473	0.2194	0.216	0.2198	0.2529	0.2529	0.2528
6	899	0.0676	0.0652	0.0679	0.2609	0.2564	0.2614	0.2602	0.2597	0.2603
7	671	0.0607	0.0627	0.0605	0.2497	0.2543	0.2492	0.2526	0.2518	0.2527
8	403	0.0606	0.0587	0.0608	0.2474	0.2434	0.2478	0.2524	0.2534	0.2522
9	197	0.0578	0.0587	0.0577	0.2416	0.2432	0.2414	0.2534	0.2526	0.2535
10	339	0.061	0.058	0.0613	0.2483	0.2415	0.249	0.253	0.2525	0.2531

Note: This table shows the R^2, Pearson correlation and Mean Absolute Error (MAE) of the prediction at the point when training stops. Due to the early stopping logic, training stops after 50 iterates in which performance on the loss function on the test set has not improved.

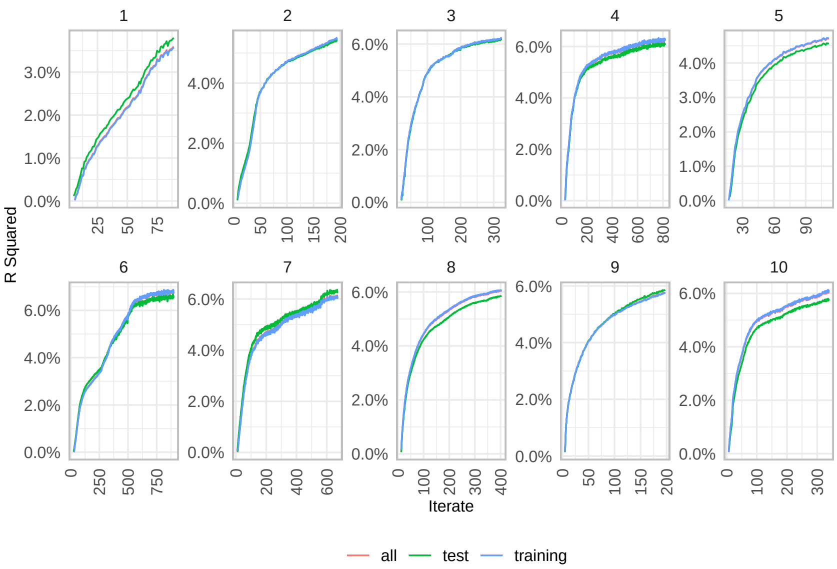
Table (J.7.2) Proportion of invalid values for each trained model.

Seed	Demand > 1	Demand < 0	Cost > Bid	Cost < 0
Combined	0%	0%	0.0006%	0
1	0%	0%	0%	0
2	0%	0%	0%	0
3	0%	0%	0.3253%	0
4	0%	0%	0.0007%	0
5	0.0004%	0.0004%	0%	0
6	0.0041%	0.0041%	0%	0
7	0.0009%	0.0009%	0.1583%	0
8	0.0002%	0.0002%	0.2376%	0
9	0%	0%	0%	0
10	0.0004%	0.0004%	0%	0

Note: This table shows fraction of demands that are outside the unit interval and the fraction of costs that are above the final bid cost or below zero. Each figure in this table is a percentage of the non-missing values in the dataset. Figures are expressed to 4 decimal places.

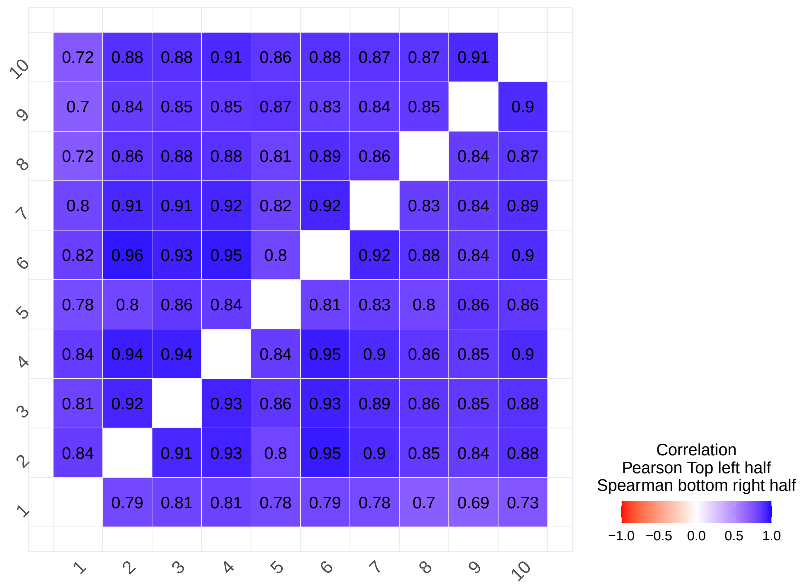
We first present tables and figures summarising the convergence path and fit of the model. The convergence plot of each seed with this specification can be seen in Figure J.7.1, while the correlation matrix between implied costs generated with each seed is displayed in Figure J.7.2. We summarise the final convergences achieved by each seed in Table J.7.1. The fraction of demands and costs that fall outside the feasible range is in Table J.7.2. Note that models that do not converge well may have more estimated probabilities outside the unit interval. Our analysis only uses the 5 trained models that have the highest R^2 however these poorly performing cases are not used. The variable importance of each of the model's input variables is summarised in Table J.7.3. We then present kernel densities of demand, cost, margin, o_1, o_2 and o_3 in Figure J.7.3. In Table J.7.4 (corresponding to Table 4) we provide the regression coefficients of margin with different fixed effects for winners and losers.

Figure (J.7.1) Convergence Plot for Specification 7



Note: This chart plots the convergence of each trained model in terms of the R^2. We only plot positive R^2 for each which is why some seeds appear to start later on. In general, R^2 values start training at negative values.

Figure (J.7.2) Correlation Matrix Between Costs Implied by Each Seed for Specification 7



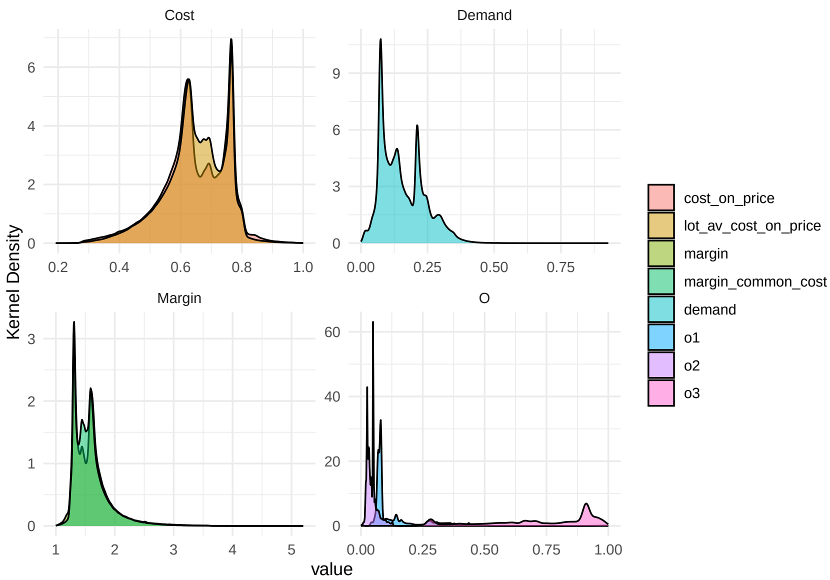
Note: This chart plots the correlation between $\frac{\text{Implied cost}}{\text{final price bid}}$ using the implied costs from each trained model. The top left half of the matrix are Pearson correlations while the bottom right half are Spearman correlations.

Table (J.7.3) Variable Importance Plot

Permuted Variable	Relative MSE	Relative Spearman Correlation
Price	9.419	0.6626
Best Preceding Price	12.6735	0.7268
Log Expected Price	1.0056	0.9516
Num. Bidders after	1.0487	0.5255
Mean prices bidder	1.0594	0.5287
Num. Participants	1.0299	0.7868
Prop. Inactive Bidders	1.0305	0.7698
Price Reduction in Round 3	1.0019	0.9845
Round 2 std of bids	1.1259	0.4379
Year tender published	1.0023	0.9801

Note: To construct this table we took the 5 trained models (of 10 in total) with above median performance on the test set. We recorded it's performance on the full (test + training) sample as a benchmark. We then permuted each of the input variables (30 times per model) and recorded their performance. We recorded the relative MSE and Spearman correlation for each permutation and each model. We then took the medians for each model. Finally we took the median across models to get one relative performance metric per permuted variable. A relative MSE above 1 indicates a deterioration in performance by permuting that variable while a relative spearman below 1 indicates the same.

Figure (J.7.3) Kernel Densities of Values Estimated with Specification 7



Note: This chart plots the kernel densities of demand, implied margin, cost and the o_1, o_2, o_3 parameters. For the purposes of this figure, we remove observations where $\frac{\text{Implied cost}}{\text{final price bid}}$ is outside the $[0, 1]$ range or demand is outside this same range. The amount of observations that are dropped in this way are small and represented in Table J.7.2.

Table (J.7.4) The Year-End Effect for Margins (with Varying Fixed Effects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Winners								
Last month	2.63***	2.56***	2.04***	1.97***	2.14***	2.09***	1.92***	2.01***
	(0.448)	(0.439)	(0.431)	(0.402)	(0.431)	(0.402)	(0.427)	(0.427)
R^2	0.00143	0.0315	0.3	0.142	0.302	0.164	0.403	0.405
N	136756	136751	136756	136756	136751	136751	136756	136751
Panel B: Losers								
Last month	3.93***	3.65***	3.14***	3.23***	3.24***	3.19***	2.95***	3.04***
	(0.587)	(0.573)	(0.521)	(0.506)	(0.521)	(0.505)	(0.491)	(0.49)
R^2	0.00213	0.037	0.232	0.126	0.235	0.151	0.32	0.323
N	382559	382550	382559	382559	382550	382550	382559	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	No	Yes	No	No	Yes	Yes	No	Yes
Seller FE	No	No	Yes	No	Yes	No	Yes	Yes
Buyer FE	No	No	No	Yes	No	Yes	Yes	Yes

Note: The sample is composed of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is a margin common cost as defined in equation 12. Last month is a dummy variable that takes the value one if the tender was finalised in December or November and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and by seller. *p<0.1; **p<0.05; ***p<0.01.

The regressions with the interaction terms are in Table J.7.5 (which corresponds to Table 5).

Finally in Table J.7.5 we replicate the regressions without controls on different time intervals of data. Specifically, for all bids, winners and losers respectively we perform separate regressions: one on 2017-2019 (pre-Covid period) and one on 2020-2021 (where the bidding process was impacted by covid). We present our findings in Table J.7.6.

Table (J.7.5) The Year-End Effect for Margins (with Varying Controls)

	All bids			*Winners*			*Losers*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Last month	2.59***	1.8***	1.73***	2.01***	1.26***	1.6***	3.04***	2.17***	1.94***
	(0.431)	(0.404)	(0.499)	(0.427)	(0.41)	(0.549)	(0.49)	(0.457)	(0.572)
Last month × 4 bidders			0.472			-0.0353			0.727
			(0.686)			(0.814)			(0.737)
Last month × 5 bidders			1.32			0.332			1.64*
			(0.926)			(1.16)			(0.977)
Last month × 6 bidders			-1.14			-2.37*			-0.745
			(1.11)			(1.32)			(1.19)
Last month × 7+ bidders			-1.4			-2.45**			-1.36
			(1.04)			(1.24)			(1.11)
4 bidders		-1.28***	-1.35***		-3.59***	-3.59***		-1.02***	-1.14***
		(0.253)	(0.272)		(0.284)	(0.308)		(0.277)	(0.298)
5 bidders		-4.35***	-4.54***		-8.21***	-8.25***		-4.21***	-4.45***
		(0.323)	(0.333)		(0.378)	(0.402)		(0.344)	(0.357)
6 bidders		-8.56***	-8.41***		-13.0***	-12.7***		-8.48***	-8.39***
		(0.408)	(0.435)		(0.486)	(0.523)		(0.437)	(0.464)
7+ bidders		-16.0***	-15.8***		-18.1***	-17.8***		-16.7***	-16.6***
		(0.421)	(0.435)		(0.506)	(0.518)		(0.451)	(0.466)
R^2	0.292	0.304	0.304	0.405	0.42	0.42	0.323	0.336	0.336
N	519301	519301	519301	136751	136751	136751	382550	382550	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Columns 1, 4, and 7 correspond to specification 13. Columns 2, 5, and 8 also include dummies for each number of bidders in the auction. Columns 3, 6, and 9 further split the last month dummy by the number of bidders in the auction. Year FE stands for year fixed effects, buyer FE refers to government department fixed effects, seller FE indicates firm fixed effects, and sector FE represents sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

Table (J.7.6) The Year-End Effect for Margins (with Split by Year Interval)

	All 17-19	*All 20-21*	*Winners 17-19*	*Winners 20-21*	*Losers 17-19*	*Losers 20-21*
	(1)	(2)	(3)	(4)	(5)	(6)
Last month	2.5***	2.26***	2.34***	1.49***	2.78***	2.66***
	(0.55)	(0.56)	(0.656)	(0.577)	(0.637)	(0.638)
R^2	0.376	0.343	0.5	0.49	0.416	0.376
N	208764	310537	56309	80442	152455	230095
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation 12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

J.8 Specification 8

This is the benchmark specification but with a cost floor of 0.3 rather than 0.25.

Table (J.8.1) Final convergence

		R^2			Pearson Corr			MAE		
Seed	Num. Iterates	All	Test	Training	All	Test	Training	All	Test	Training
1	111	0.0382	0.0406	0.038	0.1971	0.2036	0.1963	0.2579	0.2577	0.2579
2	1081	0.0689	0.0682	0.069	0.2632	0.2621	0.2634	0.2526	0.2539	0.2525
3	813	0.0666	0.0663	0.0666	0.2592	0.2584	0.2593	0.2589	0.2593	0.2589
4	886	0.0653	0.0631	0.0656	0.2576	0.253	0.2581	0.2523	0.2525	0.2523
5	209	0.0542	0.0527	0.0543	0.2339	0.2307	0.2343	0.2532	0.2532	0.2532
6	428	0.0608	0.0593	0.0609	0.2466	0.2438	0.2469	0.2569	0.2563	0.257
7	504	0.0531	0.0547	0.0529	0.2448	0.2493	0.2443	0.2441	0.2434	0.2442
8	347	0.0588	0.0579	0.0589	0.2444	0.2426	0.2446	0.2511	0.2521	0.251
9	224	0.0597	0.0608	0.0596	0.2464	0.2484	0.2462	0.251	0.2501	0.2511
10	685	0.0619	0.0591	0.0622	0.2502	0.2439	0.2509	0.2522	0.2517	0.2523

Note: This table shows the R^2, Pearson correlation and Mean Absolute Error (MAE) of the prediction at the point when training stops. Due to the early stopping logic, training stops after 50 iterates in which performance on the loss function on the test set has not improved.

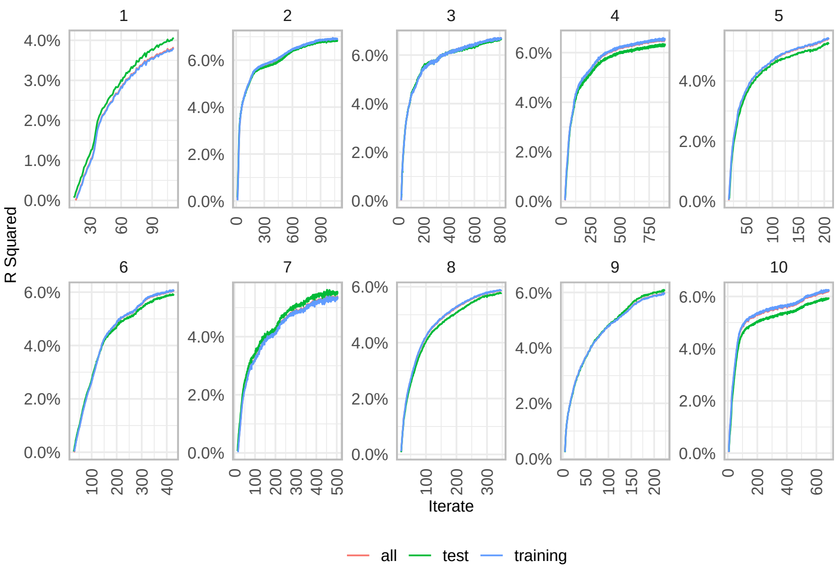
Table (J.8.2) Proportion of invalid values for each trained model.

Seed	Demand > 1	Demand < 0	Cost > Bid	Cost < 0
Combined	0.0014%	0.0014%	0.0329%	0
1	0.0004%	0.0004%	0%	0
2	0.0009%	0.0009%	0.0259%	0
3	0.0018%	0.0018%	0.0479%	0
4	0.0025%	0.0025%	0.0936%	0
5	0.0011%	0.0011%	0%	0
6	0.0041%	0.0041%	0.2247%	0
7	0.0002%	0.0002%	0.0226%	0
8	0.0002%	0.0002%	0%	0
9	0.0009%	0.0009%	0%	0
10	0.0002%	0.0002%	0.0512%	0

Note: This table shows fraction of demands that are outside the unit interval and the fraction of costs that are above the final bid cost or below zero. Each figure in this table is a percentage of the non-missing values in the dataset. Figures are expressed to 4 decimal places.

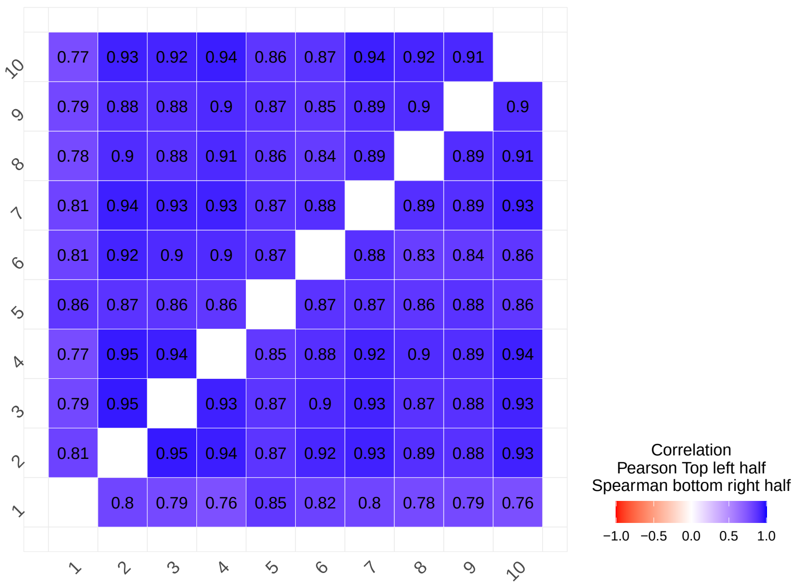
We first present tables and figures summarising the convergence path and fit of the model. The convergence plot of each seed with this specification can be seen in Figure J.8.1, while the correlation matrix between implied costs generated with each seed is displayed in Figure J.8.2. We summarise the final convergences achieved by each seed in Table J.8.1. The fraction of demands and costs that fall outside the feasible range is in Table J.8.2. Note that models that do not converge well may have more estimated probabilities outside the unit interval. Our analysis only uses the 5 trained models that have the highest R^2 however these poorly performing cases are not used. The variable importance of each of the model's input variables is summarised in Table J.8.3. We then present kernel densities of demand, cost, margin, o_1, o_2 and o_3 in Figure J.8.3. In Table J.8.4 (corresponding to Table 4) we provide the regression coefficients of margin with different fixed effects for winners and losers.

Figure (J.8.1)   Convergence Plot for Specification 8



Note: This chart plots the convergence of each trained model in terms of the R^2. We only plot positive R^2 for each which is why some seeds appear to start later on. In general, R^2 values start training at negative values.

Figure (J.8.2)   Correlation Matrix Between Costs Implied by Each Seed for Specification 8



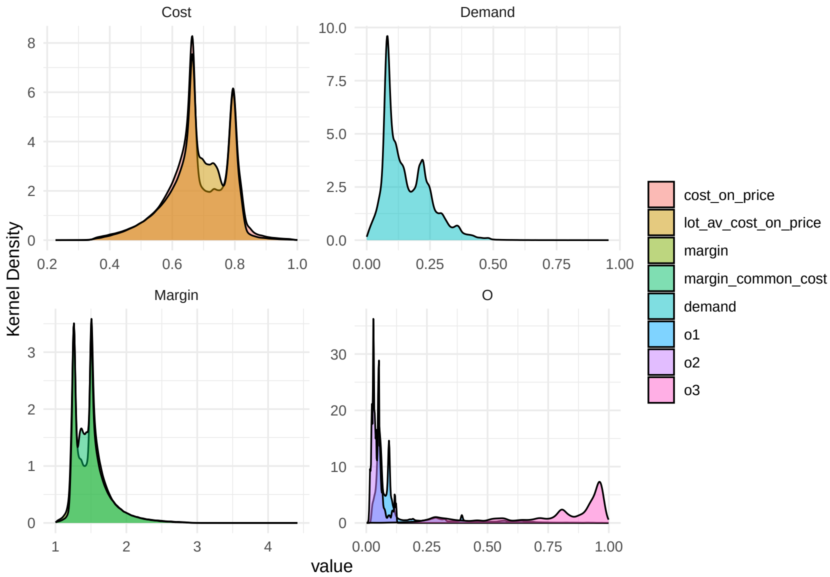
Note: This chart plots the correlation between $\frac{\text{Implied cost}}{\text{final price bid}}$ using the implied costs from each trained model. The top left half of the matrix are Pearson correlations while the bottom right half are Spearman correlations.

Table (J.8.3) Variable Importance Plot

Permuted Variable	Relative MSE	Relative Spearman Correlation
Price	27.6196	0.6438
Best Preceding Price	651.6681	0.7131
Log Expected Price	1.0052	0.9616
Num. Bidders after	1.0614	0.5137
Mean prices bidder	1.0669	0.5547
Num. Participants	1.0343	0.8173
Prop. Inactive Bidders	1.0309	0.8156
Price Reduction in Round 3	1.0024	0.9823
Round 2 std of bids	1.1404	0.4291
Year tender published	1.0025	0.9813

Note: To construct this table we took the 5 trained models (of 10 in total) with above median performance on the test set. We recorded it's performance on the full (test + training) sample as a benchmark. We then permuted each of the input variables (30 times per model) and recorded their performance. We recorded the relative MSE and Spearman correlation for each permutation and each model. We then took the medians for each model. Finally we took the median across models to get one relative performance metric per permuted variable. A relative MSE above 1 indicates a deterioration in performance by permuting that variable while a relative spearman below 1 indicates the same.

Figure (J.8.3) Kernel Densities of Values Estimated with Specification 8



Note: This chart plots the kernel densities of demand, implied margin, cost and the o_1, o_2, o_3 parameters. For the purposes of this figure, we remove observations where $\frac{\text{Implied cost}}{\text{final price bid}}$ is outside the $[0, 1]$ range or demand is outside this same range. The amount of observations that are dropped in this way are small and represented in Table J.8.2.

Table (J.8.4) The Year-End Effect for Margins (with Varying Fixed Effects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Winners								
Last month	1.91***	1.9***	1.53***	1.44***	1.61***	1.57***	1.47***	1.54***
	(0.349)	(0.344)	(0.339)	(0.317)	(0.339)	(0.318)	(0.34)	(0.34)
R^2	0.00144	0.0301	0.3	0.142	0.302	0.163	0.404	0.406
N	136756	136751	136756	136756	136751	136751	136756	136751
Panel B: Losers								
Last month	3.16***	2.98***	2.54***	2.64***	2.62***	2.64***	2.42***	2.5***
	(0.463)	(0.455)	(0.41)	(0.403)	(0.41)	(0.404)	(0.389)	(0.389)
R^2	0.00201	0.0355	0.231	0.124	0.234	0.147	0.318	0.321
N	382559	382550	382559	382559	382550	382550	382559	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	No	Yes	No	No	Yes	Yes	No	Yes
Seller FE	No	No	Yes	No	Yes	No	Yes	Yes
Buyer FE	No	No	No	Yes	No	Yes	Yes	Yes

Note: The sample is composed of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is a margin common cost as defined in equation 12. Last month is a dummy variable that takes the value one if the tender was finalised in December or November and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and by seller. *p<0.1; **p<0.05; ***p<0.01.

The regressions with the interaction terms are in Table J.8.5 (which corresponds to Table 5).

Finally in Table J.8.5 we replicate the regressions without controls on different time intervals of data. Specifically, for all bids, winners and losers respectively we perform separate regressions: one on 2017-2019 (pre-Covid period) and one on 2020-2021 (where the bidding process was impacted by covid). We present our findings in Table J.8.6.

Table (J.8.5) The Year-End Effect for Margins (with Varying Controls)

	All bids			*Winners*			*Losers*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Last month	2.09***	1.48***	1.25***	1.54***	0.943***	1.09**	2.5***	1.82***	1.44***
	(0.341)	(0.32)	(0.382)	(0.34)	(0.326)	(0.431)	(0.389)	(0.363)	(0.438)
Last month × 4 bidders			0.595			0.2			0.8
			(0.548)			(0.655)			(0.591)
Last month × 5 bidders			1.24*			0.307			1.55**
			(0.743)			(0.922)			(0.786)
Last month × 6 bidders			-0.73			-1.71			-0.385
			(0.933)			(1.1)			(0.993)
Last month × 7+ bidders			-0.515			-1.46			-0.462
			(0.894)			(1.04)			(0.958)
4 bidders		0.0848	-0.00929		-2.11***	-2.14***		0.35	0.222
		(0.204)	(0.218)		(0.229)	(0.248)		(0.226)	(0.242)
5 bidders		-2.01***	-2.19***		-5.74***	-5.78***		-1.83***	-2.06***
		(0.268)	(0.277)		(0.312)	(0.333)		(0.287)	(0.298)
6 bidders		-5.24***	-5.14***		-9.57***	-9.33***		-5.07***	-5.03***
		(0.34)	(0.364)		(0.405)	(0.435)		(0.365)	(0.389)
7+ bidders		-12.6***	-12.6***		-14.8***	-14.6***		-13.2***	-13.2***
		(0.353)	(0.367)		(0.416)	(0.428)		(0.38)	(0.394)
R^2	0.289	0.3	0.3	0.406	0.42	0.42	0.321	0.333	0.333
N	519301	519301	519301	136751	136751	136751	382550	382550	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Columns 1, 4, and 7 correspond to specification 13. Columns 2, 5, and 8 also include dummies for each number of bidders in the auction. Columns 3, 6, and 9 further split the last month dummy by the number of bidders in the auction. Year FE stands for year fixed effects, buyer FE refers to government department fixed effects, seller FE indicates firm fixed effects, and sector FE represents sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

Table (J.8.6) The Year-End Effect for Margins (with Split by Year Interval)

	All 17-19	*All 20-21*	*Winners 17-19*	*Winners 20-21*	*Losers 17-19*	*Losers 20-21*
	(1)	(2)	(3)	(4)	(5)	(6)
Last month	1.99***	1.9***	1.69***	1.22***	2.26***	2.26***
	(0.44)	(0.453)	(0.525)	(0.468)	(0.514)	(0.519)
R^2	0.371	0.34	0.501	0.49	0.413	0.375
N	208764	310537	56309	80442	152455	230095
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation 12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

J.9 Specification 9

This is the benchmark specification but with a cost floor of 0.4 rather than 0.25.

Table (J.9.1) Final convergence

		R^2			Pearson Corr			MAE		
Seed	Num. Iterates	All	Test	Training	All	Test	Training	All	Test	Training
1	241	0.0446	0.0473	0.0443	0.2133	0.2198	0.2126	0.2544	0.2542	0.2544
2	1081	0.0669	0.0664	0.067	0.2598	0.2593	0.2598	0.2511	0.2523	0.251
3	1042	0.0659	0.0647	0.0661	0.2576	0.255	0.2578	0.2584	0.2589	0.2583
4	413	0.0531	0.0512	0.0533	0.2313	0.2274	0.2318	0.2539	0.2542	0.2538
5	635	0.0581	0.0567	0.0582	0.2423	0.2393	0.2427	0.2522	0.2523	0.2522
6	433	0.0595	0.0586	0.0596	0.2444	0.2424	0.2447	0.2547	0.254	0.2548
7	383	0.0473	0.0484	0.0471	0.2399	0.2429	0.2395	0.2384	0.2377	0.2385
8	743	0.0584	0.0574	0.0585	0.2449	0.2429	0.2451	0.2481	0.2492	0.248
9	429	0.0616	0.0621	0.0616	0.2493	0.25	0.2492	0.2518	0.2511	0.2519
10	780	0.0633	0.0612	0.0635	0.2539	0.2489	0.2545	0.2506	0.2499	0.2507

Note: This table shows the R^2, Pearson correlation and Mean Absolute Error (MAE) of the prediction at the point when training stops. Due to the early stopping logic, training stops after 50 iterates in which performance on the loss function on the test set has not improved.

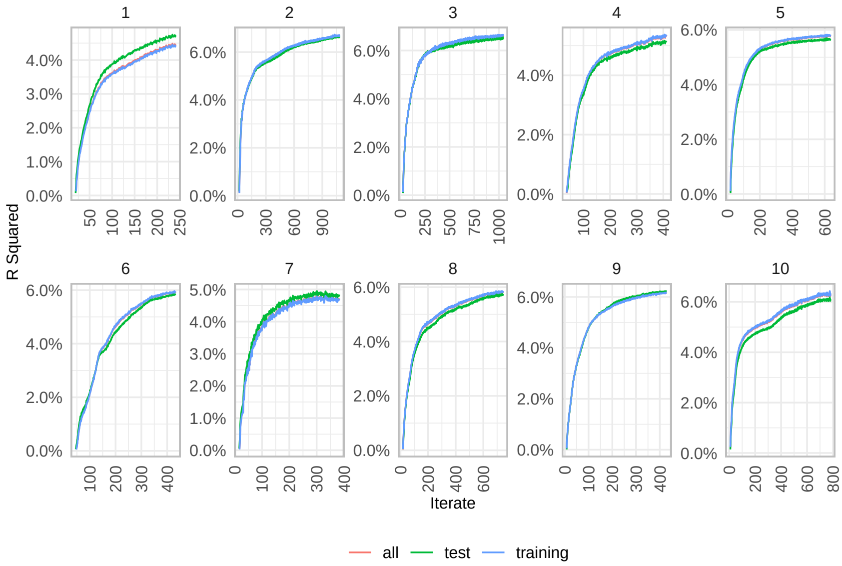
Table (J.9.2) Proportion of invalid values for each trained model.

Seed	Demand > 1	Demand < 0	Cost > Bid	Cost < 0
Combined	0.0023%	0.0023%	0.0677%	0
1	0.0006%	0.0006%	0%	0
2	0.0026%	0.0026%	0%	0
3	0.0029%	0.0029%	0.0844%	0
4	0.0021%	0.0021%	0.2555%	0
5	0.001%	0.001%	0%	0
6	0.0011%	0.0011%	0.0896%	0
7	0.0011%	0.0011%	0.0002%	0
8	0.0006%	0.0006%	0.0002%	0
9	0.0026%	0.0026%	0.9363%	0
10	0.0028%	0.0028%	0.0718%	0

Note: This table shows fraction of demands that are outside the unit interval and the fraction of costs that are above the final bid cost or below zero. Each figure in this table is a percentage of the non-missing values in the dataset. Figures are expressed to 4 decimal places.

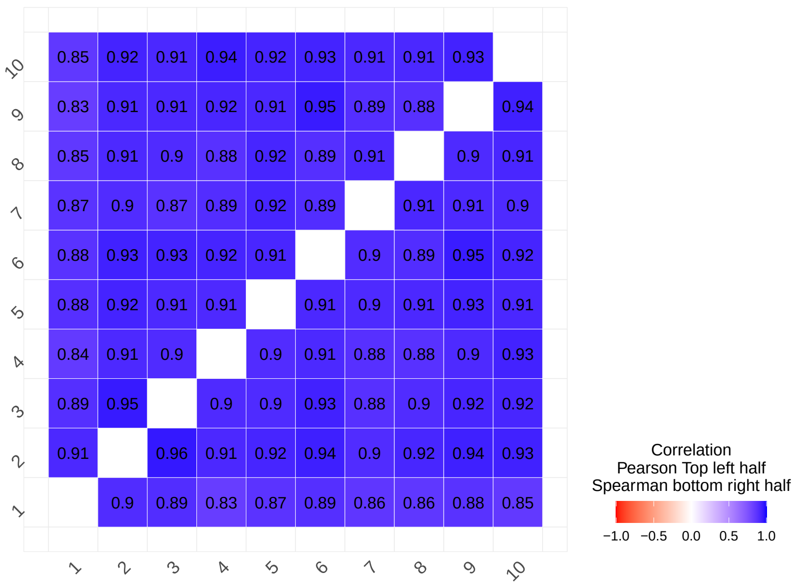
We first present tables and figures summarising the convergence path and fit of the model. The convergence plot of each seed with this specification can be seen in Figure J.9.1, while the correlation matrix between implied costs generated with each seed is displayed in Figure J.9.2. We summarise the final convergences achieved by each seed in Table J.9.1. The fraction of demands and costs that fall outside the feasible range is in Table J.9.2. Note that models that do not converge well may have more estimated probabilities outside the unit interval. Our analysis only uses the 5 trained models that have the highest R^2 however these poorly performing cases are not used. The variable importance of each of the model's input variables is summarised in Table J.9.3. We then present kernel densities of demand, cost, margin, o_1, o_2 and o_3 in Figure J.9.3. In Table J.9.4 (corresponding to Table 4) we provide the regression coefficients of margin with different fixed effects for winners and losers.

Figure (J.9.1) Convergence Plot for Specification 9



Note: This chart plots the convergence of each trained model in terms of the R^2. We only plot positive R^2 for each which is why some seeds appear to start later on. In general, R^2 values start training at negative values.

Figure (J.9.2) Correlation Matrix Between Costs Implied by Each Seed for Specification 9



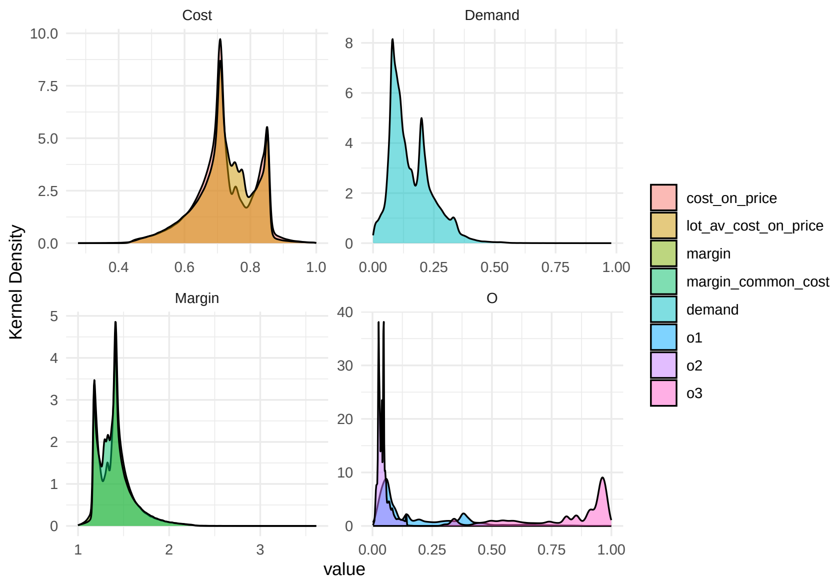
Note: This chart plots the correlation between $\frac{\text{Implied cost}}{\text{final price bid}}$ using the implied costs from each trained model. The top left half of the matrix are Pearson correlations while the bottom right half are Spearman correlations.

Table (J.9.3) Variable Importance Plot

Permuted Variable	Relative MSE	Relative Spearman Correlation
Price	361.9827	0.561
Best Preceding Price	2162.3849	0.65
Log Expected Price	1.0068	0.9453
Num. Bidders after	1.0552	0.5228
Mean prices bidder	1.069	0.5424
Num. Participants	1.0411	0.8034
Prop. Inactive Bidders	1.0492	0.709
Price Reduction in Round 3	1.0016	0.9865
Round 2 std of bids	1.1566	0.4264
Year tender published	1.0027	0.9824

Note: To construct this table we took the 5 trained models (of 10 in total) with above median performance on the test set. We recorded it's performance on the full (test + training) sample as a benchmark. We then permuted each of the input variables (30 times per model) and recorded their performance. We recorded the relative MSE and Spearman correlation for each permutation and each model. We then took the medians for each model. Finally we took the median across models to get one relative performance metric per permuted variable. A relative MSE above 1 indicates a deterioration in performance by permuting that variable while a relative spearman below 1 indicates the same.

Figure (J.9.3) Kernel Densities of Values Estimated with Specification 9



Note: This chart plots the kernel densities of demand, implied margin, cost and the o_1, o_2, o_3 parameters. For the purposes of this figure, we remove observations where $\frac{\text{Implied cost}}{\text{final price bid}}$ is outside the $[0, 1]$ range or demand is outside this same range. The amount of observations that are dropped in this way are small and represented in Table J.9.2.

Table (J.9.4) The Year-End Effect for Margins (with Varying Fixed Effects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Winners								
Last month	1.15***	1.15***	0.955***	0.793***	1.01***	0.901***	0.899***	0.95***
	(0.25)	(0.247)	(0.246)	(0.232)	(0.246)	(0.233)	(0.25)	(0.25)
R^2	0.00121	0.0274	0.301	0.141	0.302	0.16	0.404	0.406
N	136756	136751	136756	136756	136751	136751	136756	136751
Panel B: Losers								
Last month	2.13***	1.98***	1.63***	1.7***	1.69***	1.72***	1.53***	1.59***
	(0.331)	(0.327)	(0.294)	(0.29)	(0.294)	(0.292)	(0.282)	(0.282)
R^2	0.00243	0.0344	0.232	0.117	0.235	0.139	0.314	0.317
N	382559	382550	382559	382559	382550	382550	382559	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	No	Yes	No	No	Yes	Yes	No	Yes
Seller FE	No	No	Yes	No	Yes	No	Yes	Yes
Buyer FE	No	No	No	Yes	No	Yes	Yes	Yes

Note: The sample is composed of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is a margin common cost as defined in equation 12. Last month is a dummy variable that takes the value one if the tender was finalised in December or November and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and by seller. *p<0.1; **p<0.05; ***p<0.01.

The regressions with the interaction terms are in Table J.9.5 (which corresponds to Table 5).

Finally in Table J.9.5 we replicate the regressions without controls on different time intervals of data. Specifically, for all bids, winners and losers respectively we perform separate regressions: one on 2017-2019 (pre-Covid period) and one on 2020-2021 (where the bidding process was impacted by covid). We present our findings in Table J.9.6.

Table (J.9.5) The Year-End Effect for Margins (with Varying Controls)

	All bids			*Winners*			*Losers*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Last month	1.29***	1.02***	0.891***	0.95***	0.588**	0.756**	1.59***	1.29***	1.06***
	(0.245)	(0.236)	(0.285)	(0.25)	(0.243)	(0.323)	(0.282)	(0.27)	(0.329)
Last month × 4 bidders			0.43			0.114			0.565
			(0.409)			(0.497)			(0.445)
Last month × 5 bidders			0.802			-0.0269			1.05*
			(0.554)			(0.701)			(0.592)
Last month × 6 bidders			-0.561			-1.38			-0.297
			(0.709)			(0.838)			(0.756)
Last month × 7+ bidders			-0.551			-1.18			-0.567
			(0.716)			(0.817)			(0.775)
4 bidders		0.274*	0.207		-1.69***	-1.71***		0.525***	0.435**
		(0.154)	(0.165)		(0.175)	(0.19)		(0.172)	(0.185)
5 bidders		-0.917***	-1.04***		-4.36***	-4.36***		-0.701***	-0.856***
		(0.2)	(0.206)		(0.239)	(0.255)		(0.216)	(0.224)
6 bidders		-2.1***	-2.03***		-6.34***	-6.15***		-1.86***	-1.82***
		(0.256)	(0.275)		(0.309)	(0.331)		(0.276)	(0.295)
7+ bidders		-5.51***	-5.45***		-8.6***	-8.46***		-5.78***	-5.73***
		(0.27)	(0.281)		(0.326)	(0.335)		(0.292)	(0.302)
R^2	0.282	0.286	0.286	0.406	0.415	0.415	0.317	0.322	0.322
N	519301	519301	519301	136751	136751	136751	382550	382550	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Columns 1, 4, and 7 correspond to specification 13. Columns 2, 5, and 8 also include dummies for each number of bidders in the auction. Columns 3, 6, and 9 further split the last month dummy by the number of bidders in the auction. Year FE stands for year fixed effects, buyer FE refers to government department fixed effects, seller FE indicates firm fixed effects, and sector FE represents sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

Table (J.9.6) The Year-End Effect for Margins (with Split by Year Interval)

	All 17-19	*All 20-21*	*Winners 17-19*	*Winners 20-21*	*Losers 17-19*	*Losers 20-21*
	(1)	(2)	(3)	(4)	(5)	(6)
Last month	1.41***	1.03***	1.2***	0.668*	1.64***	1.26***
	(0.33)	(0.334)	(0.399)	(0.348)	(0.39)	(0.386)
R^2	0.362	0.333	0.499	0.493	0.406	0.37
N	208764	310537	56309	80442	152455	230095
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation 12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

J.10 Specification 10

This is the benchmark specification but with a cost floor of 0.5 rather than 0.25.

Table (J.10.1) Final convergence

		R^2			Pearson Corr			MAE		
Seed	Num. Iterates	All	Test	Training	All	Test	Training	All	Test	Training
1	1127	0.0625	0.0656	0.0622	0.2522	0.2587	0.2515	0.2492	0.2489	0.2492
2	884	0.0612	0.0605	0.0612	0.2485	0.2477	0.2485	0.2509	0.2522	0.2507
3	233	0.0574	0.0568	0.0574	0.2409	0.24	0.241	0.2521	0.2526	0.2521
4	743	0.0548	0.0521	0.0551	0.2397	0.2334	0.2404	0.2484	0.2488	0.2483
5	1040	0.0554	0.0537	0.0556	0.2378	0.2339	0.2382	0.2507	0.2508	0.2507
6	499	0.0483	0.0459	0.0486	0.2203	0.2154	0.2209	0.2612	0.2606	0.2612
7	776	0.0538	0.0541	0.0537	0.2387	0.2393	0.2387	0.2468	0.2462	0.2468
8	327	0.0528	0.0517	0.0529	0.232	0.2298	0.2322	0.2521	0.253	0.2519
9	959	0.0657	0.0668	0.0656	0.2566	0.2585	0.2564	0.2526	0.2518	0.2527
10	1059	0.0621	0.0598	0.0624	0.2509	0.2456	0.2514	0.2514	0.2507	0.2515

Note: This table shows the R^2, Pearson correlation and Mean Absolute Error (MAE) of the prediction at the point when training stops. Due to the early stopping logic, training stops after 50 iterates in which performance on the loss function on the test set has not improved.

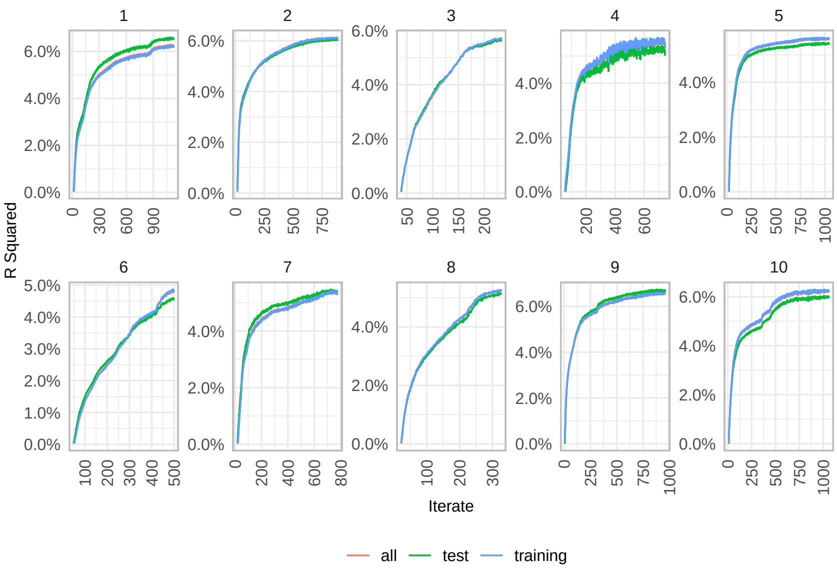
Table (J.10.2) Proportion of invalid values for each trained model.

Seed	Demand > 1	Demand < 0	Cost > Bid	Cost < 0
Combined	0.0045%	0.0045%	0.0012%	0
1	0.0049%	0.0049%	0%	0
2	0.0059%	0.0059%	0%	0
3	0.0086%	0.0086%	0.6899%	0
4	0.0047%	0.0047%	0.0381%	0
5	0.0042%	0.0042%	0%	0
6	0.0058%	0.0058%	0.1988%	0
7	0.0038%	0.0038%	0%	0
8	0.0048%	0.0048%	0.1834%	0
9	0.0104%	0.0104%	0.0377%	0
10	0.0104%	0.0104%	0.1391%	0

Note: This table shows fraction of demands that are outside the unit interval and the fraction of costs that are above the final bid cost or below zero. Each figure in this table is a percentage of the non-missing values in the dataset. Figures are expressed to 4 decimal places.

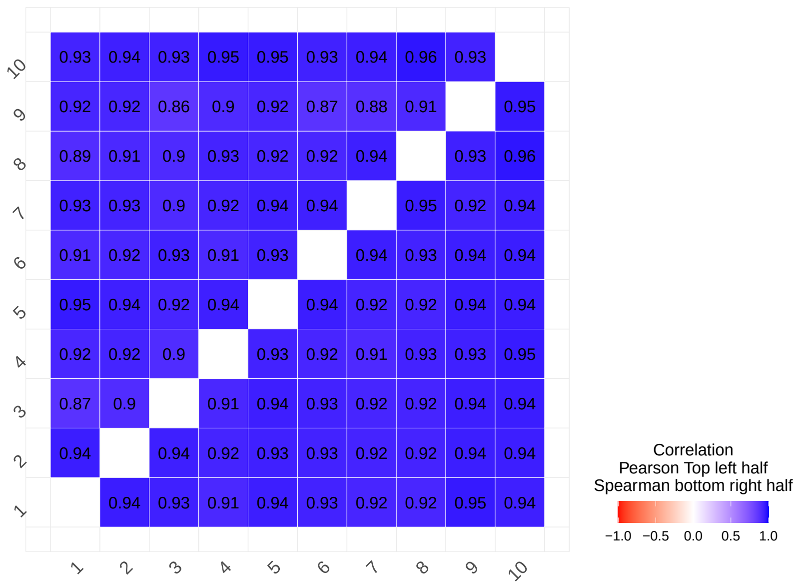
We first present tables and figures summarising the convergence path and fit of the model. The convergence plot of each seed with this specification can be seen in Figure J.10.1, while the correlation matrix between implied costs generated with each seed is displayed in Figure J.10.2. We summarise the final convergences achieved by each seed in Table J.10.1. The fraction of demands and costs that fall outside the feasible range is in Table J.10.2. Note that models that do not converge well may have more estimated probabilities outside the unit interval. Our analysis only uses the 5 trained models that have the highest R^2 however these poorly performing cases are not used. The variable importance of each of the model's input variables is summarised in Table J.10.3. We then present kernel densities of demand, cost, margin, o_1, o_2 and o_3 in Figure J.10.3. In Table J.10.4 (corresponding to Table 4) we provide the regression coefficients of margin with different fixed effects for winners and losers.

Figure (J.10.1) Convergence Plot for Specification 10



Note: This chart plots the convergence of each trained model in terms of the R^2. We only plot positive R^2 for each which is why some seeds appear to start later on. In general, R^2 values start training at negative values.

Figure (J.10.2) Correlation Matrix Between Costs Implied by Each Seed for Specification 10



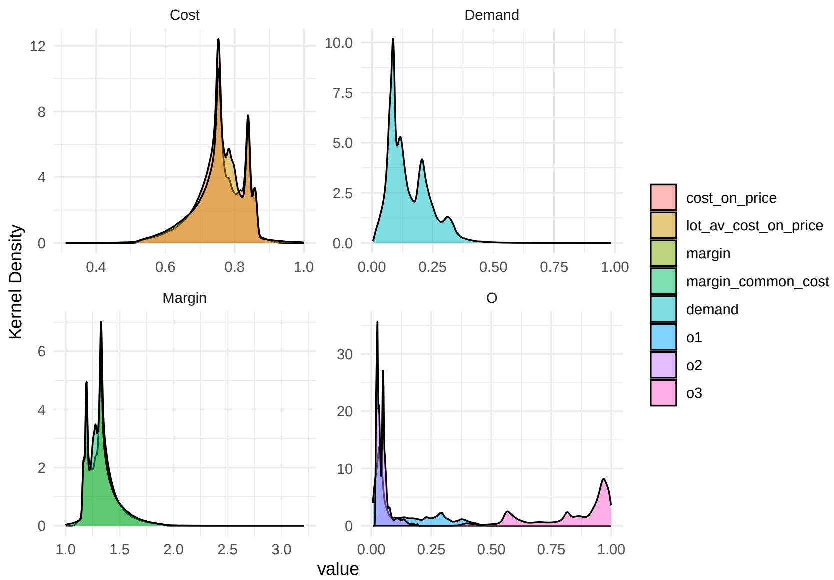
Note: This chart plots the correlation between $\frac{\text{Implied cost}}{\text{final price bid}}$ using the implied costs from each trained model. The top left half of the matrix are Pearson correlations while the bottom right half are Spearman correlations.

Table (J.10.3) Variable Importance Plot

Permuted Variable	Relative MSE	Relative Spearman Correlation
Price	12197.9501	0.5363
Best Preceding Price	31639.2783	0.6322
Log Expected Price	1.0051	0.9579
Num. Bidders after	1.0547	0.5253
Mean prices bidder	1.0833	0.5356
Num. Participants	1.0497	0.7062
Prop. Inactive Bidders	1.0679	0.6521
Price Reduction in Round 3	1.0017	0.9906
Round 2 std of bids	1.1693	0.4067
Year tender published	1.0029	0.9819

Note: To construct this table we took the 5 trained models (of 10 in total) with above median performance on the test set. We recorded it's performance on the full (test + training) sample as a benchmark. We then permuted each of the input variables (30 times per model) and recorded their performance. We recorded the relative MSE and Spearman correlation for each permutation and each model. We then took the medians for each model. Finally we took the median across models to get one relative performance metric per permuted variable. A relative MSE above 1 indicates a deterioration in performance by permuting that variable while a relative spearman below 1 indicates the same.

Figure (J.10.3) Kernel Densities of Values Estimated with Specification 10



Note: This chart plots the kernel densities of demand, implied margin, cost and the o_1, o_2, o_3 parameters. For the purposes of this figure, we remove observations where $\frac{\text{Implied cost}}{\text{final price bid}}$ is outside the $[0, 1]$ range or demand is outside this same range. The amount of observations that are dropped in this way are small and represented in Table J.10.2.

Table (J.10.4) The Year-End Effect for Margins (with Varying Fixed Effects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Winners								
Last month	0.802***	0.769***	0.677***	0.58***	0.718***	0.62***	0.65***	0.689***
	(0.189)	(0.184)	(0.186)	(0.176)	(0.187)	(0.175)	(0.189)	(0.189)
R^2	0.000777	0.0205	0.296	0.14	0.297	0.154	0.4	0.402
N	136756	136751	136756	136756	136751	136751	136756	136751
Panel B: Losers								
Last month	1.67***	1.48***	1.17***	1.37***	1.21***	1.32***	1.14***	1.18***
	(0.256)	(0.243)	(0.216)	(0.222)	(0.216)	(0.219)	(0.211)	(0.21)
R^2	0.00276	0.0268	0.227	0.104	0.229	0.12	0.303	0.305
N	382559	382550	382559	382559	382550	382550	382559	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	No	Yes	No	No	Yes	Yes	No	Yes
Seller FE	No	No	Yes	No	Yes	No	Yes	Yes
Buyer FE	No	No	No	Yes	No	Yes	Yes	Yes

Note: The sample is composed of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is a margin common cost as defined in equation 12. Last month is a dummy variable that takes the value one if the tender was finalised in December or November and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and by seller. *p<0.1; **p<0.05; ***p<0.01.

The regressions with the interaction terms are in Table J.10.5 (which corresponds to Table 5).

Finally in Table J.10.5 we replicate the regressions without controls on different time intervals of data. Specifically, for all bids, winners and losers respectively we perform separate regressions: one on 2017-2019 (pre-Covid period) and one on 2020-2021 (where the bidding process was impacted by covid). We present our findings in Table J.10.6.

Table (J.10.5) The Year-End Effect for Margins (with Varying Controls)

	All bids			*Winners*			*Losers*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Last month	0.939***	0.764***	0.599***	0.689***	0.392**	0.488*	1.18***	0.99***	0.75***
	(0.181)	(0.176)	(0.216)	(0.189)	(0.184)	(0.252)	(0.21)	(0.204)	(0.252)
Last month × 4 bidders			0.377			0.148			0.459
			(0.318)			(0.394)			(0.35)
Last month × 5 bidders			0.641			-0.104			0.839*
			(0.425)			(0.553)			(0.457)
Last month × 6 bidders			-0.2			-0.979			0.0269
			(0.543)			(0.65)			(0.58)
Last month × 7+ bidders			-0.172			-0.637			-0.232
			(0.513)			(0.624)			(0.554)
4 bidders		0.284**	0.224*		-1.47***	-1.49***		0.505***	0.432***
		(0.122)	(0.13)		(0.14)	(0.153)		(0.138)	(0.148)
5 bidders		-0.306*	-0.401**		-3.5***	-3.49***		-0.0764	-0.202
		(0.156)	(0.161)		(0.191)	(0.206)		(0.169)	(0.176)
6 bidders		-1.27***	-1.24***		-5.21***	-5.08***		-1.02***	-1.03***
		(0.196)	(0.209)		(0.245)	(0.262)		(0.212)	(0.226)
7+ bidders		-3.6***	-3.59***		-7.11***	-7.04***		-3.69***	-3.68***
		(0.194)	(0.204)		(0.241)	(0.25)		(0.211)	(0.222)
R^2	0.268	0.271	0.271	0.402	0.412	0.412	0.305	0.308	0.308
N	519301	519301	519301	136751	136751	136751	382550	382550	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Columns 1, 4, and 7 correspond to specification 13. Columns 2, 5, and 8 also include dummies for each number of bidders in the auction. Columns 3, 6, and 9 further split the last month dummy by the number of bidders in the auction. Year FE stands for year fixed effects, buyer FE refers to government department fixed effects, seller FE indicates firm fixed effects, and sector FE represents sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

Table (J.10.6) The Year-End Effect for Margins (with Split by Year Interval)

	All 17-19	*All 20-21*	*Winners 17-19*	*Winners 20-21*	*Losers 17-19*	*Losers 20-21*
	(1)	(2)	(3)	(4)	(5)	(6)
Last month	1.21***	0.628**	0.938***	0.426	1.45***	0.808***
	(0.25)	(0.247)	(0.313)	(0.268)	(0.3)	(0.288)
R^2	0.347	0.316	0.494	0.491	0.394	0.355
N	208764	310537	56309	80442	152455	230095
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation 12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

J.11 Specification 11

This is the benchmark specification but with a cost floor of 0.625 rather than 0.25.

Table (J.11.1) Final convergence

		R^2			Pearson Corr			MAE		
Seed	Num. Iterates	All	Test	Training	All	Test	Training	All	Test	Training
1	1391	0.0519	0.056	0.0514	0.231	0.24	0.2301	0.2491	0.2486	0.2491
2	824	0.0551	0.0551	0.0551	0.2391	0.24	0.239	0.2473	0.2485	0.2472
3	563	0.0487	0.0476	0.0488	0.2246	0.2229	0.2248	0.2494	0.25	0.2494
4	435	0.0453	0.0422	0.0456	0.2175	0.2114	0.2181	0.2501	0.2505	0.25
5	150	-0.0031	-0.0051	-0.0029	0.1009	0.0955	0.1014	0.2538	0.2537	0.2538
6	457	0.0386	0.0361	0.0389	0.2022	0.1968	0.2028	0.2592	0.2586	0.2593
7	429	0.0335	0.0342	0.0334	0.1886	0.1902	0.1884	0.2529	0.2524	0.253
8	1177	0.05	0.0484	0.0502	0.2296	0.2263	0.23	0.2461	0.2472	0.246
9	1213	0.0569	0.0574	0.0568	0.2409	0.2417	0.2408	0.2499	0.249	0.25
10	240	0.0357	0.0333	0.0359	0.1966	0.1902	0.1974	0.2497	0.2491	0.2497

Note: This table shows the R^2, Pearson correlation and Mean Absolute Error (MAE) of the prediction at the point when training stops. Due to the early stopping logic, training stops after 50 iterates in which performance on the loss function on the test set has not improved.

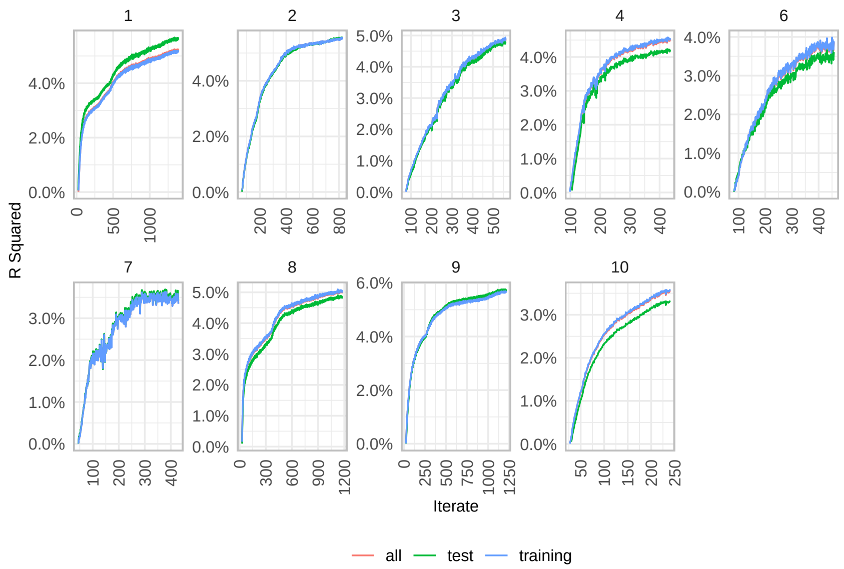
Table (J.11.2) Proportion of invalid values for each trained model.

Seed	Demand > 1	Demand < 0	Cost > Bid	Cost < 0
Combined	0.0262%	0.0262%	0.007%	0
1	0.0217%	0.0217%	0%	0
2	0.0356%	0.0356%	0.2329%	0
3	0.0235%	0.0235%	0.0527%	0
4	0.0264%	0.0264%	0.3998%	0
5	0.0192%	0.0192%	1.5997%	0
6	0.0371%	0.0371%	0.7623%	0
7	0.0165%	0.0165%	0.137%	0
8	0.0216%	0.0216%	0.009%	0
9	0.0296%	0.0296%	0%	0
10	0.0152%	0.0152%	0.6599%	0

Note: This table shows fraction of demands that are outside the unit interval and the fraction of costs that are above the final bid cost or below zero. Each figure in this table is a percentage of the non-missing values in the dataset. Figures are expressed to 4 decimal places.

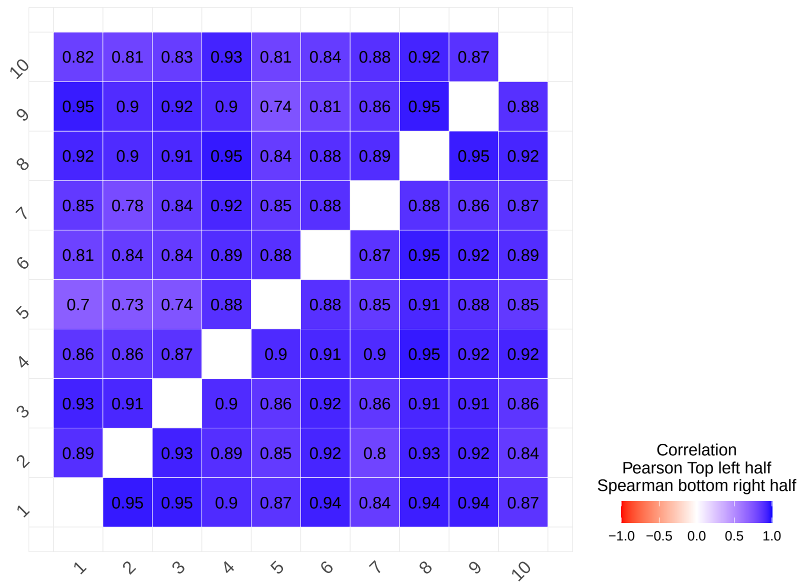
We first present tables and figures summarising the convergence path and fit of the model. The convergence plot of each seed with this specification can be seen in Figure J.11.1, while the correlation matrix between implied costs generated with each seed is displayed in Figure J.11.2. We summarise the final convergences achieved by each seed in Table J.11.1. The fraction of demands and costs that fall outside the feasible range is in Table J.11.2. Note that models that do not converge well may have more estimated probabilities outside the unit interval. Our analysis only uses the 5 trained models that have the highest R^2 however these poorly performing cases are not used. The variable importance of each of the model's input variables is summarised in Table J.11.3. We then present kernel densities of demand, cost, margin, o_1, o_2 and o_3 in Figure J.11.3. In Table J.11.4 (corresponding to Table 4) we provide the regression coefficients of margin with different fixed effects for winners and losers.

Figure (J.11.1) Convergence Plot for Specification 11



Note: This chart plots the convergence of each trained model in terms of the R^2. We only plot positive R^2 for each which is why some seeds appear to start later on. In general, R^2 values start training at negative values.

Figure (J.11.2) Correlation Matrix Between Costs Implied by Each Seed for Specification 11



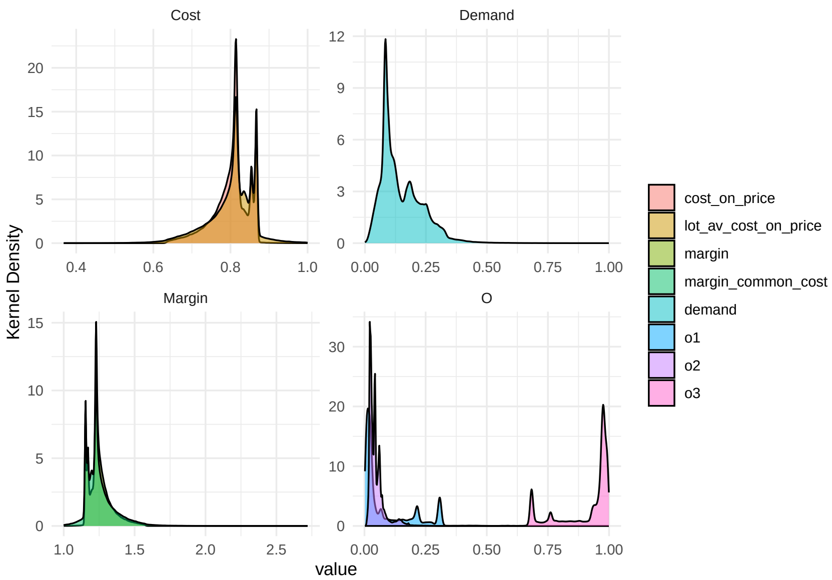
Note: This chart plots the correlation between $\frac{\text{Implied cost}}{\text{final price bid}}$ using the implied costs from each trained model. The top left half of the matrix are Pearson correlations while the bottom right half are Spearman correlations.

Table (J.11.3) Variable Importance Plot

Permuted Variable	Relative MSE	Relative Spearman Correlation
Price	15461255.3092	0.4519
Best Preceding Price	10084617.6252	0.5732
Log Expected Price	1.0048	0.9609
Num. Bidders after	1.0474	0.5481
Mean prices bidder	1.0672	0.5163
Num. Participants	1.0508	0.7697
Prop. Inactive Bidders	1.0326	0.8059
Price Reduction in Round 3	1.0028	0.9855
Round 2 std of bids	1.2214	0.358
Year tender published	1.0024	0.9846

Note: To construct this table we took the 5 trained models (of 10 in total) with above median performance on the test set. We recorded it's performance on the full (test + training) sample as a benchmark. We then permuted each of the input variables (30 times per model) and recorded their performance. We recorded the relative MSE and Spearman correlation for each permutation and each model. We then took the medians for each model. Finally we took the median across models to get one relative performance metric per permuted variable. A relative MSE above 1 indicates a deterioration in performance by permuting that variable while a relative spearman below 1 indicates the same.

Figure (J.11.3) Kernel Densities of Values Estimated with Specification 11



Note: This chart plots the kernel densities of demand, implied margin, cost and the o_1, o_2, o_3 parameters. For the purposes of this figure, we remove observations where $\frac{\text{Implied cost}}{\text{final price bid}}$ is outside the $[0, 1]$ range or demand is outside this same range. The amount of observations that are dropped in this way are small and represented in Table J.11.2.

Table (J.11.4) The Year-End Effect for Margins (with Varying Fixed Effects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Winners								
Last month	0.293**	0.251*	0.254*	0.155	0.28**	0.165	0.23*	0.255*
	(0.137)	(0.132)	(0.135)	(0.13)	(0.136)	(0.128)	(0.138)	(0.139)
R^2	0.000483	0.0135	0.296	0.14	0.297	0.149	0.401	0.402
N	136756	136751	136756	136756	136751	136751	136756	136751
Panel B: Losers								
Last month	0.977***	0.78***	0.509***	0.753***	0.534***	0.685***	0.513***	0.537***
	(0.179)	(0.164)	(0.144)	(0.154)	(0.144)	(0.149)	(0.145)	(0.145)
R^2	0.00366	0.0204	0.225	0.0891	0.227	0.1	0.293	0.295
N	382559	382550	382559	382559	382550	382550	382559	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	No	Yes	No	No	Yes	Yes	No	Yes
Seller FE	No	No	Yes	No	Yes	No	Yes	Yes
Buyer FE	No	No	No	Yes	No	Yes	Yes	Yes

Note: The sample is composed of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is a margin common cost as defined in equation 12. Last month is a dummy variable that takes the value one if the tender was finalised in December or November and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and by seller. *p<0.1; **p<0.05; ***p<0.01.

The regressions with the interaction terms are in Table J.11.5 (which corresponds to Table 5).

Finally in Table J.11.5 we replicate the regressions without controls on different time intervals of data. Specifically, for all bids, winners and losers respectively we perform separate regressions: one on 2017-2019 (pre-Covid period) and one on 2020-2021 (where the bidding process was impacted by covid). We present our findings in Table J.11.6.

Table (J.11.5) The Year-End Effect for Margins (with Varying Controls)

	All bids (1)	(2)	(3)	*Winners* (4)	(5)	(6)	*Losers* (7)	(8)	(9)
Last month	0.365***	0.45***	0.303*	0.255*	0.135	0.222	0.537***	0.636***	0.434**
	(0.123)	(0.124)	(0.159)	(0.139)	(0.138)	(0.192)	(0.145)	(0.145)	(0.188)
Last month × 4 bidders			0.265			0.0865			0.308
			(0.237)			(0.303)			(0.265)
Last month × 5 bidders			0.451			-0.178			0.607*
			(0.311)			(0.423)			(0.34)
Last month × 6 bidders			-0.0274			-0.715			0.143
			(0.399)			(0.494)			(0.428)
Last month × 7+ bidders			0.0577			-0.435			0.0106
			(0.375)			(0.471)			(0.406)
4 bidders		0.771***	0.729***		-0.839***	-0.853***		0.992***	0.943***
		(0.0922)	(0.0985)		(0.11)	(0.119)		(0.106)	(0.113)
5 bidders		1.19***	1.12***		-1.86***	-1.84***		1.46***	1.37***
		(0.116)	(0.12)		(0.148)	(0.161)		(0.127)	(0.132)
6 bidders		1.45***	1.45***		-2.51***	-2.41***		1.79***	1.77***
		(0.145)	(0.154)		(0.186)	(0.199)		(0.158)	(0.168)
7+ bidders		1.7***	1.69***		-2.67***	-2.62***		1.91***	1.9***
		(0.141)	(0.147)		(0.175)	(0.183)		(0.152)	(0.16)
R^2	0.255	0.256	0.256	0.402	0.405	0.405	0.295	0.296	0.296
N	519301	519301	519301	136751	136751	136751	382550	382550	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Columns 1, 4, and 7 correspond to specification 13. Columns 2, 5, and 8 also include dummies for each number of bidders in the auction. Columns 3, 6, and 9 further split the last month dummy by the number of bidders in the auction. Year FE stands for year fixed effects, buyer FE refers to government department fixed effects, seller FE indicates firm fixed effects, and sector FE represents sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

Table (J.11.6) The Year-End Effect for Margins (with Split by Year Interval)

	All 17-19 (1)	*All 20-21* (2)	*Winners 17-19* (3)	*Winners 20-21* (4)	*Losers 17-19* (5)	*Losers 20-21* (6)
Last month	0.825***	-0.0308	0.556**	-0.00358	1.04***	0.0657
	(0.182)	(0.174)	(0.238)	(0.199)	(0.223)	(0.203)
R^2	0.331	0.3	0.491	0.494	0.382	0.342
N	208764	310537	56309	80442	152455	230095
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation 12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

J.12 Specification 12

This is the benchmark specification but with a cost floor of 0.75 rather than 0.25.

Table (J.12.1) Final convergence

		R^2			Pearson Corr			MAE		
Seed	Num. Iterates	All	Test	Training	All	Test	Training	All	Test	Training
1	738	0.0134	0.0191	0.0128	0.162	0.1735	0.1607	0.2406	0.2402	0.2406
2	1017	0.0174	0.0181	0.0174	0.1739	0.1772	0.1735	0.2474	0.2486	0.2473
3	937	0.0124	0.0096	0.0127	0.173	0.1686	0.1735	0.2325	0.2333	0.2325
4	820	0.0224	0.017	0.023	0.1828	0.1741	0.1838	0.2435	0.2441	0.2435
5	941	0.0175	0.0155	0.0178	0.175	0.1709	0.1755	0.2415	0.2417	0.2415
6	660	0.0171	0.0179	0.0171	0.1773	0.1764	0.1774	0.2393	0.2384	0.2394
7	865	0.0195	0.0173	0.0197	0.1797	0.1775	0.1799	0.2394	0.2388	0.2395
8	275	-0.0353	-0.0392	-0.0348	0.0714	0.0659	0.0721	0.2399	0.2409	0.2398
9	1038	0.0244	0.0228	0.0245	0.19	0.1874	0.1903	0.239	0.2385	0.2391
10	284	0.0079	0.0096	0.0077	0.1539	0.1526	0.1541	0.2394	0.2386	0.2395

Note: This table shows the R^2, Pearson correlation and Mean Absolute Error (MAE) of the prediction at the point when training stops. Due to the early stopping logic, training stops after 50 iterates in which performance on the loss function on the test set has not improved.

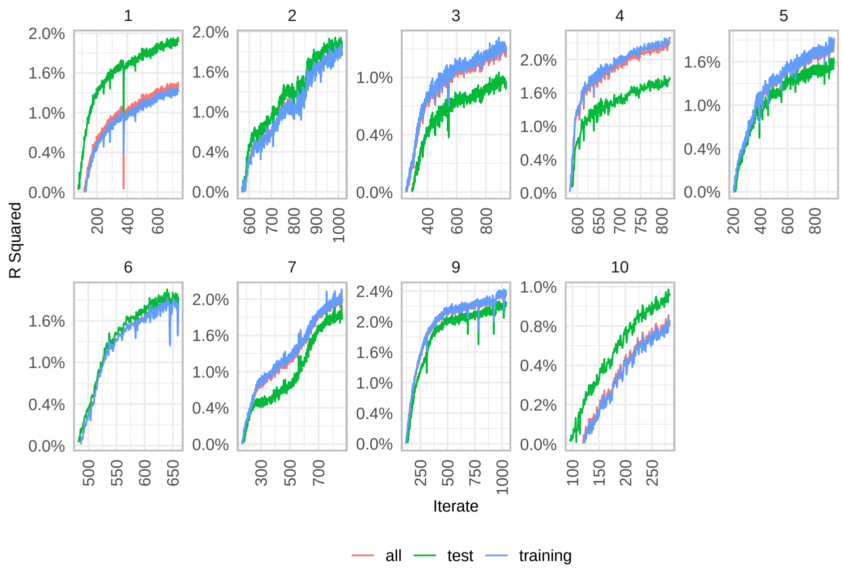
Table (J.12.2) Proportion of invalid values for each trained model.

Seed	Demand > 1	Demand < 0	Cost > Bid	Cost < 0
Combined	0.0678%	0.0678%	0%	0
1	0.0527%	0.0527%	0%	0
2	0.0776%	0.0776%	0%	0
3	0.0415%	0.0415%	0.9547%	0
4	0.0788%	0.0788%	0.0131%	0
5	0.0648%	0.0648%	0.5161%	0
6	0.0738%	0.0738%	1.2992%	0
7	0.0716%	0.0716%	0%	0
8	0.0774%	0.0774%	0.4934%	0
9	0.0674%	0.0674%	0.0346%	0
10	0.0566%	0.0566%	0.7439%	0

Note: This table shows fraction of demands that are outside the unit interval and the fraction of costs that are above the final bid cost or below zero. Each figure in this table is a percentage of the non-missing values in the dataset. Figures are expressed to 4 decimal places.

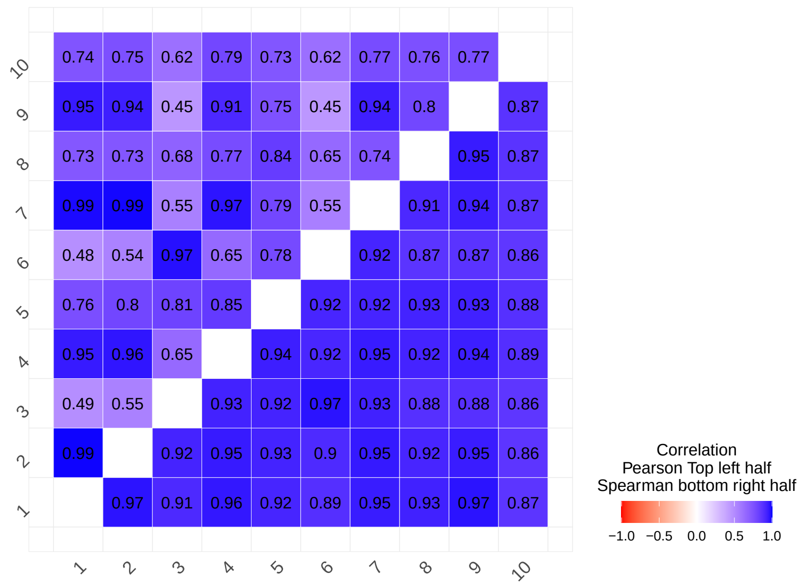
We first present tables and figures summarising the convergence path and fit of the model. The convergence plot of each seed with this specification can be seen in Figure J.12.1, while the correlation matrix between implied costs generated with each seed is displayed in Figure J.12.2. We summarise the final convergences achieved by each seed in Table J.12.1. The fraction of demands and costs that fall outside the feasible range is in Table J.12.2. Note that models that do not converge well may have more estimated probabilities outside the unit interval. Our analysis only uses the 5 trained models that have the highest R^2 however these poorly performing cases are not used. The variable importance of each of the model's input variables is summarised in Table J.12.3. We then present kernel densities of demand, cost, margin, o_1, o_2 and o_3 in Figure J.12.3. In Table J.12.4 (corresponding to Table 4) we provide the regression coefficients of margin with different fixed effects for winners and losers.

Figure (J.12.1) Convergence Plot for Specification 12



Note: This chart plots the convergence of each trained model in terms of the R^2. We only plot positive R^2 for each which is why some seeds appear to start later on. In general, R^2 values start training at negative values.

Figure (J.12.2) Correlation Matrix Between Costs Implied by Each Seed for Specification 12



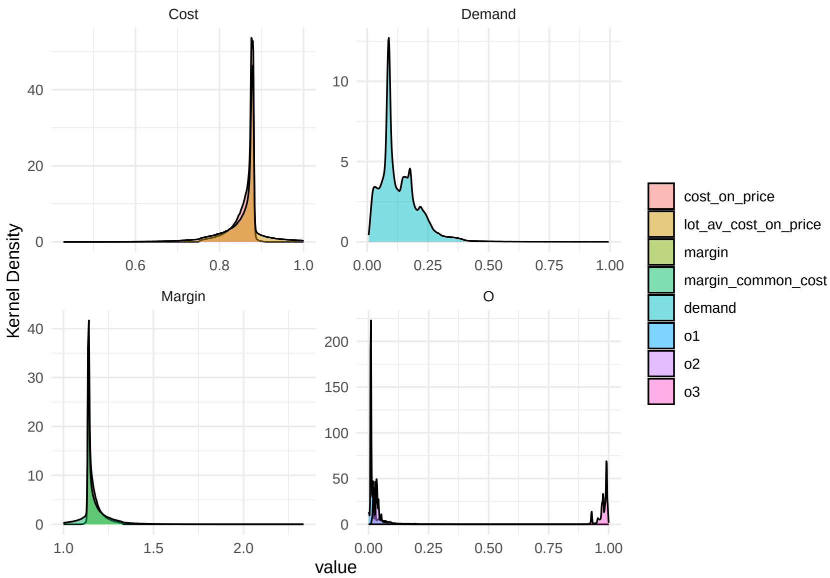
Note: This chart plots the correlation between $\frac{\text{Implied cost}}{\text{final price bid}}$ using the implied costs from each trained model. The top left half of the matrix are Pearson correlations while the bottom right half are Spearman correlations.

Table (J.12.3) Variable Importance Plot

Permuted Variable	Relative MSE	Relative Spearman Correlation
Price	27336671189951.1	0.3293
Best Preceding Price	4929355180538.38	0.5162
Log Expected Price	1.0037	0.9552
Num. Bidders after	1.0295	0.6241
Mean prices bidder	1.0977	0.443
Num. Participants	1.0307	0.7938
Prop. Inactive Bidders	1.0495	0.6223
Price Reduction in Round 3	1.0022	0.9807
Round 2 std of bids	1.2734	0.2941
Year tender published	1.0019	0.9816

Note: To construct this table we took the 5 trained models (of 10 in total) with above median performance on the test set. We recorded it's performance on the full (test + training) sample as a benchmark. We then permuted each of the input variables (30 times per model) and recorded their performance. We recorded the relative MSE and Spearman correlation for each permutation and each model. We then took the medians for each model. Finally we took the median across models to get one relative performance metric per permuted variable. A relative MSE above 1 indicates a deterioration in performance by permuting that variable while a relative spearman below 1 indicates the same.

Figure (J.12.3) Kernel Densities of Values Estimated with Specification 12



Note: This chart plots the kernel densities of demand, implied margin, cost and the o_1, o_2, o_3 parameters. For the purposes of this figure, we remove observations where $\frac{\text{Implied cost}}{\text{final price bid}}$ is outside the $[0, 1]$ range or demand is outside this same range. The amount of observations that are dropped in this way are small and represented in Table J.12.2.

Table (J.12.4) The Year-End Effect for Margins (with Varying Fixed Effects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Winners								
Last month	0.0558	0.0291	0.112	0.001	0.128	-0.0016	0.102	0.118
	(0.108)	(0.102)	(0.107)	(0.104)	(0.107)	(0.102)	(0.11)	(0.11)
R^2	0.000599	0.00872	0.3	0.141	0.301	0.146	0.405	0.405
N	136756	136751	136756	136756	136751	136751	136756	136751
Panel B: Losers								
Last month	0.646***	0.469***	0.258**	0.513***	0.272***	0.438***	0.297***	0.312***
	(0.129)	(0.115)	(0.103)	(0.113)	(0.103)	(0.108)	(0.107)	(0.107)
R^2	0.00295	0.0123	0.222	0.073	0.223	0.0791	0.282	0.283
N	382559	382550	382559	382559	382550	382550	382559	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	No	Yes	No	No	Yes	Yes	No	Yes
Seller FE	No	No	Yes	No	Yes	No	Yes	Yes
Buyer FE	No	No	No	Yes	No	Yes	Yes	Yes

Note: The sample is composed of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is a margin common cost as defined in equation 12. Last month is a dummy variable that takes the value one if the tender was finalised in December or November and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and by seller. *p<0.1; **p<0.05; ***p<0.01.

The regressions with the interaction terms are in Table J.12.5 (which corresponds to Table 5).

Finally in Table J.12.5 we replicate the regressions without controls on different time intervals of data. Specifically, for all bids, winners and losers respectively we perform separate regressions: one on 2017-2019 (pre-Covid period) and one on 2020-2021 (where the bidding process was impacted by covid). We present our findings in Table J.12.6.

Table (J.12.5) The Year-End Effect for Margins (with Varying Controls)

	All bids			*Winners*			*Losers*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Last month	0.176*	0.293***	0.181	0.118	0.0328	0.0968	0.312***	0.445***	0.3**
	(0.0912)	(0.0914)	(0.124)	(0.11)	(0.11)	(0.155)	(0.107)	(0.108)	(0.148)
Last month × 4 bidders			0.194			0.0921			0.205
			(0.183)			(0.244)			(0.209)
Last month × 5 bidders			0.296			-0.226			0.421
			(0.238)			(0.339)			(0.265)
Last month × 6 bidders			0.00673			-0.563			0.135
			(0.303)			(0.392)			(0.328)
Last month × 7+ bidders			0.11			-0.23			0.0429
			(0.285)			(0.378)			(0.31)
4 bidders		0.68***	0.649***		-0.749***	-0.763***		0.871***	0.837***
		(0.0726)	(0.0771)		(0.09)	(0.0981)		(0.0846)	(0.0896)
5 bidders		1.24***	1.2***		-1.51***	-1.47***		1.5***	1.43***
		(0.0917)	(0.094)		(0.123)	(0.134)		(0.101)	(0.105)
6 bidders		1.68***	1.68***		-1.88***	-1.8***		1.99***	1.97***
		(0.112)	(0.118)		(0.152)	(0.162)		(0.124)	(0.131)
7+ bidders		2.33***	2.32***		-1.86***	-1.83***		2.58***	2.57***
		(0.113)	(0.116)		(0.139)	(0.146)		(0.121)	(0.126)
R^2	0.242	0.245	0.245	0.405	0.408	0.408	0.283	0.286	0.286
N	519301	519301	519301	136751	136751	136751	382550	382550	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Columns 1, 4, and 7 correspond to specification 13. Columns 2, 5, and 8 also include dummies for each number of bidders in the auction. Columns 3, 6, and 9 further split the last month dummy by the number of bidders in the auction. Year FE stands for year fixed effects, buyer FE refers to government department fixed effects, seller FE indicates firm fixed effects, and sector FE represents sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

Table (J.12.6) The Year-End Effect for Margins (with Split by Year Interval)

	All 17-19	*All 20-21*	*Winners 17-19*	*Winners 20-21*	*Losers 17-19*	*Losers 20-21*
	(1)	(2)	(3)	(4)	(5)	(6)
Last month	0.626***	-0.17	0.372*	-0.0727	0.823***	-0.103
	(0.137)	(0.132)	(0.192)	(0.159)	(0.171)	(0.153)
R^2	0.317	0.286	0.492	0.499	0.371	0.328
N	208764	310537	56309	80442	152455	230095
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation 12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

J.13 Specification 13

This specification is the same as the benchmark specification but with a lower level of dropout using in the training of the neural network. Specifically there is no dropout rather than a 2.5% probability of a node dropping out.

Table (J.13.1) Final convergence

		R^2			Pearson Corr			MAE		
Seed	Num. Iterates	All	Test	Training	All	Test	Training	All	Test	Training
1	1098	0.0713	0.0744	0.0709	0.2673	0.2731	0.2667	0.2507	0.2504	0.2507
2	639	0.0708	0.0702	0.0708	0.2678	0.2669	0.2679	0.2502	0.2514	0.25
3	1276	0.0748	0.0736	0.0749	0.2738	0.2719	0.2741	0.2496	0.2501	0.2495
4	1160	0.071	0.0687	0.0713	0.267	0.2629	0.2674	0.2533	0.2536	0.2532
5	1127	0.0715	0.0696	0.0717	0.2678	0.2645	0.2682	0.252	0.2521	0.252
6	1591	0.0755	0.0737	0.0757	0.2752	0.2721	0.2756	0.2496	0.2491	0.2496
7	613	0.0643	0.0637	0.0644	0.2549	0.2536	0.2551	0.2494	0.2491	0.2494
8	268	0.0576	0.0558	0.0578	0.2408	0.2374	0.2412	0.2539	0.255	0.2538
9	555	0.0754	0.0764	0.0752	0.2748	0.2766	0.2746	0.2497	0.2488	0.2498
10	1473	0.0733	0.0708	0.0736	0.2711	0.2664	0.2716	0.25	0.2494	0.25

Note: This table shows the R^2, Pearson correlation and Mean Absolute Error (MAE) of the prediction at the point when training stops. Due to the early stopping logic, training stops after 50 iterates in which performance on the loss function on the test set has not improved.

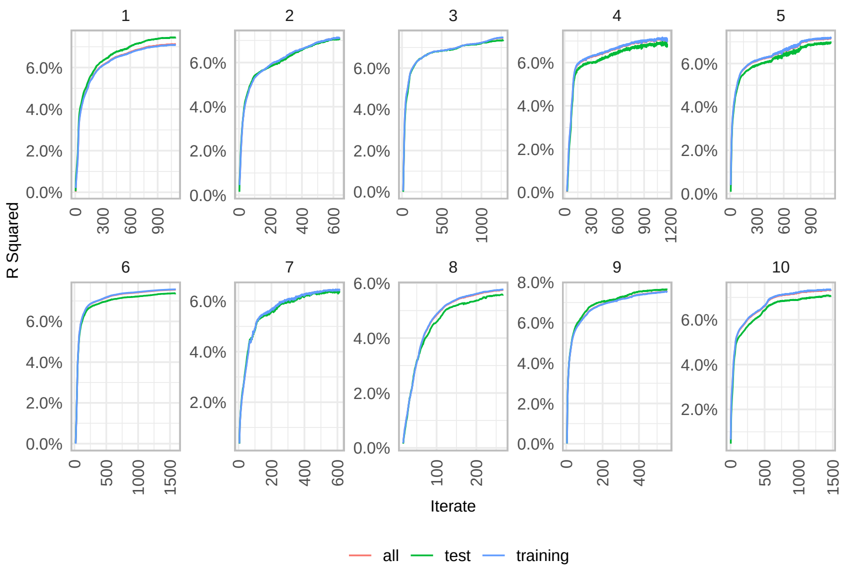
Table (J.13.2) Proportion of invalid values for each trained model.

Seed	Demand > 1	Demand < 0	Cost > Bid	Cost < 0
Combined	0.0042%	0.0042%	0.0106%	0
1	0.0041%	0.0041%	0.0005%	0
2	0.0109%	0.0109%	0%	0
3	0.0047%	0.0047%	0.0444%	0
4	0.0053%	0.0053%	0.0076%	0
5	0.0056%	0.0056%	0.0012%	0
6	0.0072%	0.0072%	0.0201%	0
7	0.0044%	0.0044%	0%	0
8	0.002%	0.002%	0.0001%	0
9	0.0093%	0.0093%	0.1201%	0
10	0.0092%	0.0092%	0%	0

Note: This table shows fraction of demands that are outside the unit interval and the fraction of costs that are above the final bid cost or below zero. Each figure in this table is a percentage of the non-missing values in the dataset. Figures are expressed to 4 decimal places.

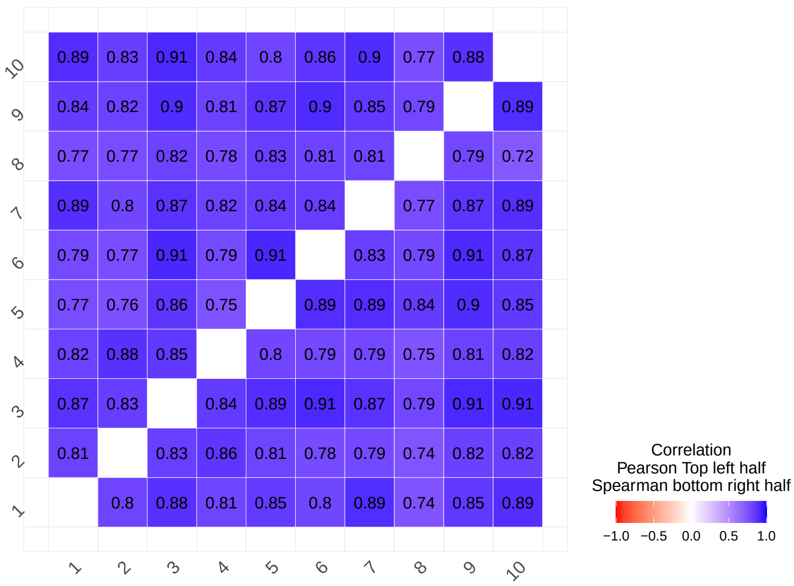
We first present tables and figures summarising the convergence path and fit of the model. The convergence plot of each seed with this specification can be seen in Figure J.13.1, while the correlation matrix between implied costs generated with each seed is displayed in Figure J.13.2. We summarise the final convergences achieved by each seed in Table J.13.1. The fraction of demands and costs that fall outside the feasible range is in Table J.13.2. Note that models that do not converge well may have more estimated probabilities outside the unit interval. Our analysis only uses the 5 trained models that have the highest R^2 however these poorly performing cases are not used. The variable importance of each of

Figure (J.13.1) Convergence Plot for Specification 13



Note: This chart plots the convergence of each trained model in terms of the R^2. We only plot positive R^2 for each which is why some seeds appear to start later on. In general, R^2 values start training at negative values.

Figure (J.13.2) Correlation Matrix Between Costs Implied by Each Seed for Specification 13



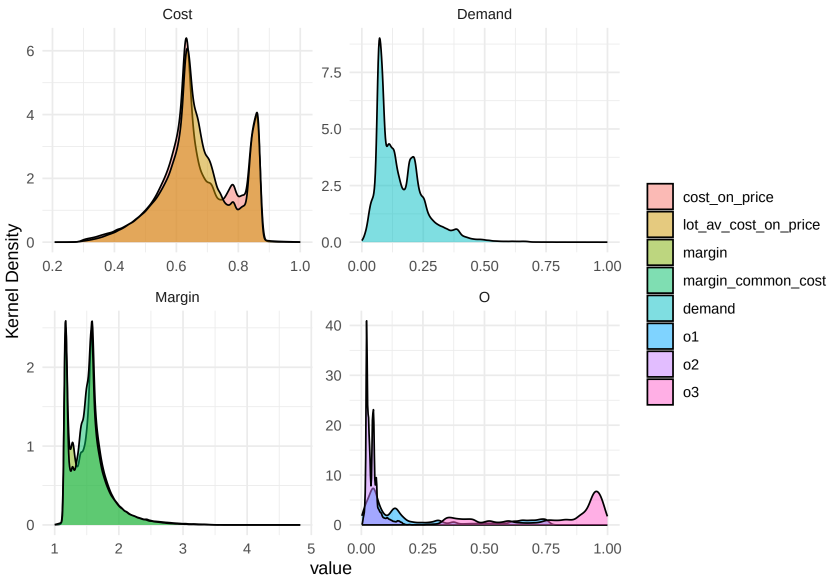
Note: This chart plots the correlation between $\frac{\text{Implied cost}}{\text{final price bid}}$ using the implied costs from each trained model. The top left half of the matrix are Pearson correlations while the bottom right half are Spearman correlations.

Table (J.13.3) Variable Importance Plot

Permuted Variable	Relative MSE	Relative Spearman Correlation
Price	11.7838	0.6636
Best Preceding Price	697.6311	0.7522
Log Expected Price	1.0091	0.9493
Num. Bidders after	1.0642	0.5262
Mean prices bidder	1.0884	0.5762
Num. Participants	1.0679	0.726
Prop. Inactive Bidders	1.0681	0.6693
Price Reduction in Round 3	1.0045	0.9771
Round 2 std of bids	1.17	0.4547
Year tender published	1.0032	0.9824

Note: To construct this table we took the 5 trained models (of 10 in total) with above median performance on the test set. We recorded it's performance on the full (test + training) sample as a benchmark. We then permuted each of the input variables (30 times per model) and recorded their performance. We recorded the relative MSE and Spearman correlation for each permutation and each model. We then took the medians for each model. Finally we took the median across models to get one relative performance metric per permuted variable. A relative MSE above 1 indicates a deterioration in performance by permuting that variable while a relative spearman below 1 indicates the same.

Figure (J.13.3) Kernel Densities of Values Estimated with Specification 13



Note: This chart plots the kernel densities of demand, implied margin, cost and the o_1, o_2, o_3 parameters. For the purposes of this figure, we remove observations where $\frac{\text{Implied cost}}{\text{final price bid}}$ is outside the $[0, 1]$ range or demand is outside this same range. The amount of observations that are dropped in this way are small and represented in Table J.13.2.

the model's input variables is summarised in Table J.13.3. We then present kernel densities of demand, cost, margin, o_1, o_2 and o_3 in Figure J.13.3. In Table J.13.4 (corresponding to Table 4) we provide the regression coefficients of margin with different fixed effects for winners and losers.

Table (J.13.4) The Year-End Effect for Margins (with Varying Fixed Effects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Winners								
Last month	2.39***	2.38***	1.94***	1.81***	2.03***	1.95***	1.84***	1.93***
	(0.426)	(0.419)	(0.413)	(0.382)	(0.412)	(0.383)	(0.407)	(0.407)
R^2	0.00132	0.0319	0.301	0.142	0.303	0.164	0.404	0.406
N	136756	136751	136756	136756	136751	136751	136756	136751
Panel B: Losers								
Last month	3.68***	3.5***	3.05***	3.03***	3.14***	3.04***	2.86***	2.95***
	(0.558)	(0.547)	(0.497)	(0.482)	(0.497)	(0.481)	(0.467)	(0.466)
R^2	0.00241	0.0392	0.234	0.126	0.237	0.152	0.322	0.325
N	382559	382550	382559	382559	382550	382550	382559	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	No	Yes	No	No	Yes	Yes	No	Yes
Seller FE	No	No	Yes	No	Yes	No	Yes	Yes
Buyer FE	No	No	No	Yes	No	Yes	Yes	Yes

Note: The sample is composed of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is a margin common cost as defined in equation 12. Last month is a dummy variable that takes the value one if the tender was finalised in December or November and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and by seller. *p<0.1; **p<0.05; ***p<0.01.

The regressions with the interaction terms are in Table J.13.5 (which corresponds to Table 5).

Finally in Table J.13.5 we replicate the regressions without controls on different time intervals of data. Specifically, for all bids, winners and losers respectively we perform separate regressions: one on 2017-2019 (pre-Covid period) and one on 2020-2021 (where the bidding process was impacted by covid). We present our findings in Table J.13.6.

Table (J.13.5) The Year-End Effect for Margins (with Varying Controls)

	All bids			*Winners*			*Losers*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Last month	2.51***	1.82***	1.7***	1.93***	1.28***	1.57***	2.95***	2.18***	1.89***
	(0.409)	(0.385)	(0.468)	(0.407)	(0.392)	(0.52)	(0.466)	(0.435)	(0.537)
Last month × 4 bidders			0.713			0.221			0.964
			(0.651)			(0.775)			(0.699)
Last month × 5 bidders			1.1			-0.0642			1.47
			(0.864)			(1.07)			(0.912)
Last month × 6 bidders			-0.916			-2.04			-0.525
			(1.04)			(1.27)			(1.11)
Last month × 7+ bidders			-1.39			-2.47**			-1.33
			(1.0)			(1.21)			(1.08)
4 bidders		-0.287	-0.399		-2.7***	-2.73***		0.0333	-0.12
		(0.239)	(0.256)		(0.266)	(0.29)		(0.264)	(0.283)
5 bidders		-4.41***	-4.57***		-8.09***	-8.08***		-4.3***	-4.52***
		(0.309)	(0.318)		(0.36)	(0.384)		(0.329)	(0.341)
6 bidders		-8.7***	-8.58***		-12.7***	-12.4***		-8.73***	-8.67***
		(0.388)	(0.413)		(0.471)	(0.503)		(0.415)	(0.439)
7+ bidders		-13.1***	-13.0***		-14.8***	-14.5***		-14.0***	-13.8***
		(0.416)	(0.428)		(0.513)	(0.525)		(0.443)	(0.454)
R^2	0.293	0.303	0.303	0.406	0.419	0.419	0.325	0.336	0.336
N	519301	519301	519301	136751	136751	136751	382550	382550	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Columns 1, 4, and 7 correspond to specification 13. Columns 2, 5, and 8 also include dummies for each number of bidders in the auction. Columns 3, 6, and 9 further split the last month dummy by the number of bidders in the auction. Year FE stands for year fixed effects, buyer FE refers to government department fixed effects, seller FE indicates firm fixed effects, and sector FE represents sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

Table (J.13.6) The Year-End Effect for Margins (with Split by Year Interval)

	All 17-19	*All 20-21*	*Winners 17-19*	*Winners 20-21*	*Losers 17-19*	*Losers 20-21*
	(1)	(2)	(3)	(4)	(5)	(6)
Last month	2.35***	2.26***	2.19***	1.44***	2.61***	2.65***
	(0.514)	(0.531)	(0.615)	(0.552)	(0.597)	(0.606)
R^2	0.375	0.344	0.5	0.491	0.416	0.378
N	208764	310537	56309	80442	152455	230095
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation 12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

J.14 Specification 14

This specification is the same as the benchmark specification but with a higher level of dropout using in the training of the neural network. Specifically there is 5% probability of a node dropping out rather than 2.5%.

Table (J.14.1) Final convergence

		R^2			Pearson Corr			MAE		
Seed	Num. Iterates	All	Test	Training	All	Test	Training	All	Test	Training
1	93	0.0252	0.0265	0.0251	0.1611	0.1658	0.1606	0.2622	0.2622	0.2622
2	178	0.0477	0.0485	0.0476	0.2193	0.2215	0.219	0.2554	0.2565	0.2553
3	210	0.0494	0.05	0.0494	0.2225	0.2239	0.2223	0.2584	0.2588	0.2584
4	219	0.0454	0.0433	0.0457	0.2152	0.2101	0.2158	0.2543	0.2546	0.2543
5	230	0.0448	0.044	0.0449	0.2122	0.2105	0.2124	0.2652	0.2651	0.2652
6	861	0.064	0.0631	0.0641	0.2541	0.2528	0.2543	0.2621	0.2614	0.2622
7	349	0.0294	0.0305	0.0293	0.221	0.2243	0.2207	0.2339	0.2333	0.234
8	646	0.0577	0.0569	0.0578	0.2415	0.2396	0.2417	0.2558	0.2568	0.2557
9	573	0.0584	0.0591	0.0583	0.242	0.2436	0.2418	0.2606	0.2598	0.2607
10	178	0.0437	0.0415	0.044	0.2111	0.2051	0.2118	0.2551	0.2545	0.2552

Note: This table shows the R^2, Pearson correlation and Mean Absolute Error (MAE) of the prediction at the point when training stops. Due to the early stopping logic, training stops after 50 iterates in which performance on the loss function on the test set has not improved.

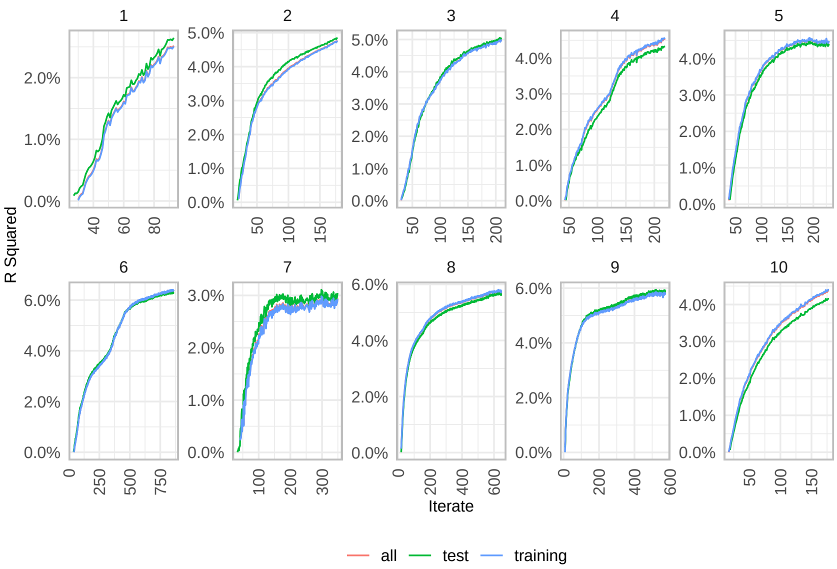
Table (J.14.2) Proportion of invalid values for each trained model.

Seed	Demand > 1	Demand < 0	Cost > Bid	Cost < 0
Combined	0.0001%	0.0001%	0.0458%	0
1	0.0001%	0.0001%	0%	0
2	0.0002%	0.0002%	0%	0
3	0%	0%	0.4171%	0
4	0.0001%	0.0001%	0%	0
5	0.0001%	0.0001%	0.3728%	0
6	0.0007%	0.0007%	0.0009%	0
7	0%	0%	0.0672%	0
8	0.0005%	0.0005%	0.2095%	0
9	0.0004%	0.0004%	0.046%	0
10	0%	0%	0.0001%	0

Note: This table shows fraction of demands that are outside the unit interval and the fraction of costs that are above the final bid cost or below zero. Each figure in this table is a percentage of the non-missing values in the dataset. Figures are expressed to 4 decimal places.

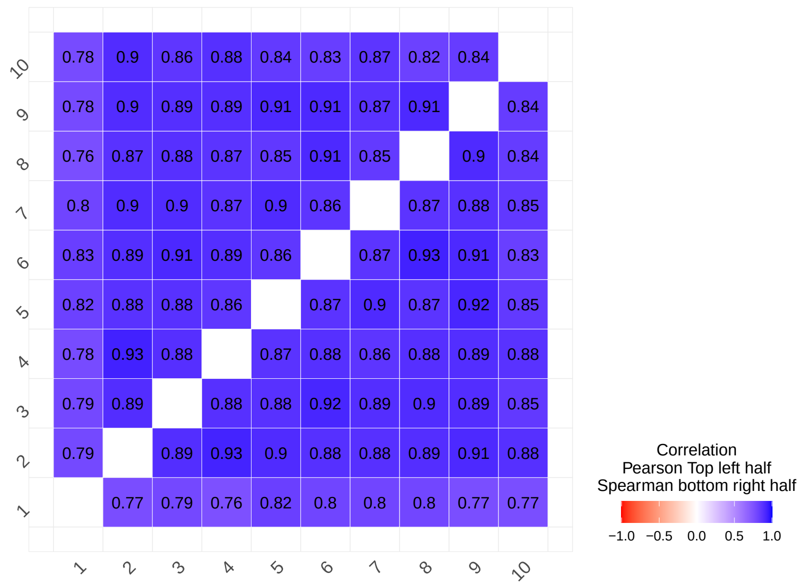
We first present tables and figures summarising the convergence path and fit of the model. The convergence plot of each seed with this specification can be seen in Figure J.14.1, while the correlation matrix between implied costs generated with each seed is displayed in Figure J.14.2. We summarise the final convergences achieved by each seed in Table J.14.1. The fraction of demands and costs that fall outside the feasible range is in Table J.14.2. Note that models that do not converge well may have more estimated probabilities outside the unit interval. Our analysis only uses the 5 trained models that have the highest R^2 however these poorly performing cases are not used. The variable importance of each of

Figure (J.14.1) Convergence Plot for Specification 14



Note: This chart plots the convergence of each trained model in terms of the R^2. We only plot positive R^2 for each which is why some seeds appear to start later on. In general, R^2 values start training at negative values.

Figure (J.14.2) Correlation Matrix Between Costs Implied by Each Seed for Specification 14



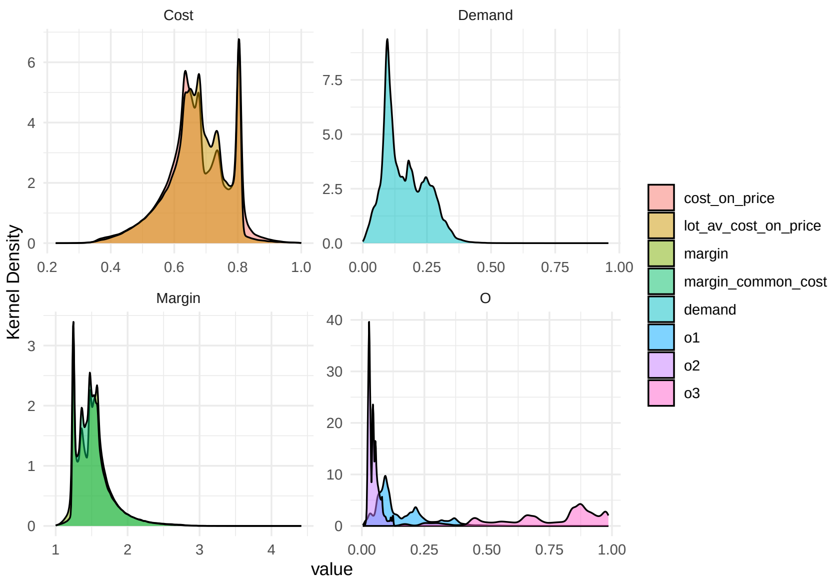
Note: This chart plots the correlation between $\frac{\text{Implied cost}}{\text{final price bid}}$ using the implied costs from each trained model. The top left half of the matrix are Pearson correlations while the bottom right half are Spearman correlations.

Table (J.14.3) Variable Importance Plot

Permuted Variable	Relative MSE	Relative Spearman Correlation
Price	9.2508	0.6121
Best Preceding Price	50.2327	0.6983
Log Expected Price	1.0045	0.9616
Num. Bidders after	1.0495	0.5095
Mean prices bidder	1.0474	0.5862
Num. Participants	1.037	0.7615
Prop. Inactive Bidders	1.0296	0.7968
Price Reduction in Round 3	1.0007	0.9952
Round 2 std of bids	1.1029	0.4716
Year tender published	1.0022	0.9818

Note: To construct this table we took the 5 trained models (of 10 in total) with above median performance on the test set. We recorded it's performance on the full (test + training) sample as a benchmark. We then permuted each of the input variables (30 times per model) and recorded their performance. We recorded the relative MSE and Spearman correlation for each permutation and each model. We then took the medians for each model. Finally we took the median across models to get one relative performance metric per permuted variable. A relative MSE above 1 indicates a deterioration in performance by permuting that variable while a relative spearman below 1 indicates the same.

Figure (J.14.3) Kernel Densities of Values Estimated with Specification 14



Note: This chart plots the kernel densities of demand, implied margin, cost and the o_1, o_2, o_3 parameters. For the purposes of this figure, we remove observations where $\frac{\text{Implied cost}}{\text{final price bid}}$ is outside the $[0, 1]$ range or demand is outside this same range. The amount of observations that are dropped in this way are small and represented in Table J.14.2.

the model's input variables is summarised in Table J.14.3. We then present kernel densities of demand, cost, margin, o_1, o_2 and o_3 in Figure J.14.3. In Table J.14.4 (corresponding to Table 4) we provide the regression coefficients of margin with different fixed effects for winners and losers.

Table (J.14.4) The Year-End Effect for Margins (with Varying Fixed Effects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Winners								
Last month	2.16***	2.04***	1.61***	1.57***	1.69***	1.66***	1.55***	1.62***
	(0.358)	(0.35)	(0.343)	(0.323)	(0.343)	(0.323)	(0.344)	(0.344)
R^2	0.00129	0.0291	0.299	0.143	0.301	0.163	0.403	0.405
N	136756	136751	136756	136756	136751	136751	136756	136751
Panel B: Losers								
Last month	3.35***	3.03***	2.59***	2.71***	2.67***	2.67***	2.48***	2.55***
	(0.467)	(0.456)	(0.416)	(0.402)	(0.415)	(0.402)	(0.393)	(0.392)
R^2	0.00249	0.036	0.232	0.124	0.235	0.147	0.319	0.322
N	382559	382550	382559	382559	382550	382550	382559	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	No	Yes	No	No	Yes	Yes	No	Yes
Seller FE	No	No	Yes	No	Yes	No	Yes	Yes
Buyer FE	No	No	No	Yes	No	Yes	Yes	Yes

Note: The sample is composed of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is a margin common cost as defined in equation 12. Last month is a dummy variable that takes the value one if the tender was finalised in December or November and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and by seller. *p<0.1; **p<0.05; ***p<0.01.

The regressions with the interaction terms are in Table J.14.5 (which corresponds to Table 5).

Finally in Table J.14.5 we replicate the regressions without controls on different time intervals of data. Specifically, for all bids, winners and losers respectively we perform separate regressions: one on 2017-2019 (pre-Covid period) and one on 2020-2021 (where the bidding process was impacted by covid). We present our findings in Table J.14.6.

Table (J.14.5) The Year-End Effect for Margins (with Varying Controls)

	All bids			*Winners*			*Losers*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Last month	2.16***	1.52***	1.44***	1.62***	0.978***	1.27***	2.55***	1.85***	1.66***
	(0.345)	(0.324)	(0.406)	(0.344)	(0.329)	(0.446)	(0.392)	(0.367)	(0.466)
Last month × 4 bidders			0.442			-0.00936			0.634
			(0.557)			(0.664)			(0.601)
Last month × 5 bidders			0.924			0.00015			1.19
			(0.75)			(0.932)			(0.794)
Last month × 6 bidders			-0.768			-1.92*			-0.427
			(0.909)			(1.08)			(0.969)
Last month × 7+ bidders			-1.05			-1.88*			-1.07
			(0.849)			(1.01)			(0.916)
4 bidders		-1.53***	-1.6***		-3.72***	-3.72***		-1.27***	-1.37***
		(0.208)	(0.223)		(0.232)	(0.252)		(0.229)	(0.245)
5 bidders		-3.52***	-3.65***		-7.22***	-7.22***		-3.34***	-3.51***
		(0.264)	(0.272)		(0.309)	(0.329)		(0.282)	(0.292)
6 bidders		-6.49***	-6.39***		-10.8***	-10.5***		-6.34***	-6.29***
		(0.332)	(0.355)		(0.399)	(0.428)		(0.356)	(0.378)
7+ bidders		-13.2***	-13.1***		-15.8***	-15.6***		-13.8***	-13.6***
		(0.336)	(0.346)		(0.41)	(0.42)		(0.361)	(0.371)
R^2	0.289	0.301	0.301	0.405	0.421	0.422	0.322	0.334	0.334
N	519301	519301	519301	136751	136751	136751	382550	382550	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Columns 1, 4, and 7 correspond to specification 13. Columns 2, 5, and 8 also include dummies for each number of bidders in the auction. Columns 3, 6, and 9 further split the last month dummy by the number of bidders in the auction. Year FE stands for year fixed effects, buyer FE refers to government department fixed effects, seller FE indicates firm fixed effects, and sector FE represents sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

Table (J.14.6) The Year-End Effect for Margins (with Split by Year Interval)

	All 17-19	*All 20-21*	*Winners 17-19*	*Winners 20-21*	*Losers 17-19*	*Losers 20-21*
	(1)	(2)	(3)	(4)	(5)	(6)
Last month	2.29***	1.79***	1.99***	1.16**	2.58***	2.13***
	(0.453)	(0.449)	(0.546)	(0.465)	(0.527)	(0.514)
R^2	0.373	0.339	0.501	0.491	0.414	0.374
N	208764	310537	56309	80442	152455	230095
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation 12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

J.15 Specification 15

This specification is the same as the benchmark specification but with only 3 hidden layers in the neural network rather than 6.

Table (J.15.1) Final convergence

		R^2			Pearson Corr			MAE		
Seed	Num. Iterates	All	Test	Training	All	Test	Training	All	Test	Training
1	611	0.0625	0.065	0.0623	0.2527	0.2585	0.2521	0.2516	0.2514	0.2516
2	788	0.0657	0.064	0.0658	0.2565	0.2531	0.2569	0.257	0.2584	0.2569
3	293	0.0626	0.0619	0.0627	0.2504	0.249	0.2505	0.2551	0.2556	0.255
4	288	0.0653	0.0641	0.0655	0.2573	0.2547	0.2576	0.2513	0.2515	0.2512
5	253	0.0552	0.0537	0.0554	0.2361	0.2328	0.2365	0.2535	0.2536	0.2535
6	151	0.0525	0.052	0.0526	0.2304	0.2289	0.2306	0.2533	0.2524	0.2533
7	1043	0.068	0.0684	0.0679	0.2608	0.2617	0.2607	0.2535	0.2529	0.2535
8	1875	0.0726	0.0718	0.0727	0.2705	0.2691	0.2707	0.2502	0.2511	0.25
9	1181	0.0719	0.0733	0.0717	0.2682	0.2708	0.2679	0.253	0.2521	0.2531
10	606	0.0674	0.0638	0.0678	0.2596	0.2526	0.2604	0.2543	0.2539	0.2544

Note: This table shows the R^2, Pearson correlation and Mean Absolute Error (MAE) of the prediction at the point when training stops. Due to the early stopping logic, training stops after 50 iterates in which performance on the loss function on the test set has not improved.

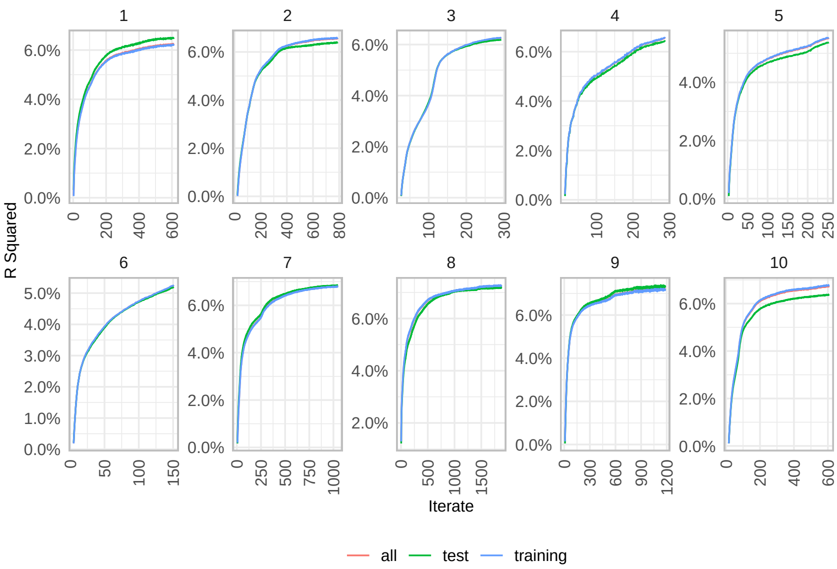
Table (J.15.2) Proportion of invalid values for each trained model.

Seed	Demand $>$ 1	Demand $<$ 0	Cost $>$ Bid	Cost $<$ 0
Combined	0.0006%	0.0006%	0.0586%	0
1	0%	0%	0.1268%	0
2	0.0005%	0.0005%	0.0417%	0
3	0.0001%	0.0001%	0.7493%	0
4	0.0006%	0.0006%	1.0264%	0
5	0.0012%	0.0012%	0.2181%	0
6	0.0015%	0.0015%	0.9923%	0
7	0%	0%	0%	0
8	0.0032%	0.0032%	0.0017%	0
9	0.0017%	0.0017%	0.0744%	0
10	0.0026%	0.0026%	0.2862%	0

Note: This table shows fraction of demands that are outside the unit interval and the fraction of costs that are above the final bid cost or below zero. Each figure in this table is a percentage of the non-missing values in the dataset. Figures are expressed to 4 decimal places.

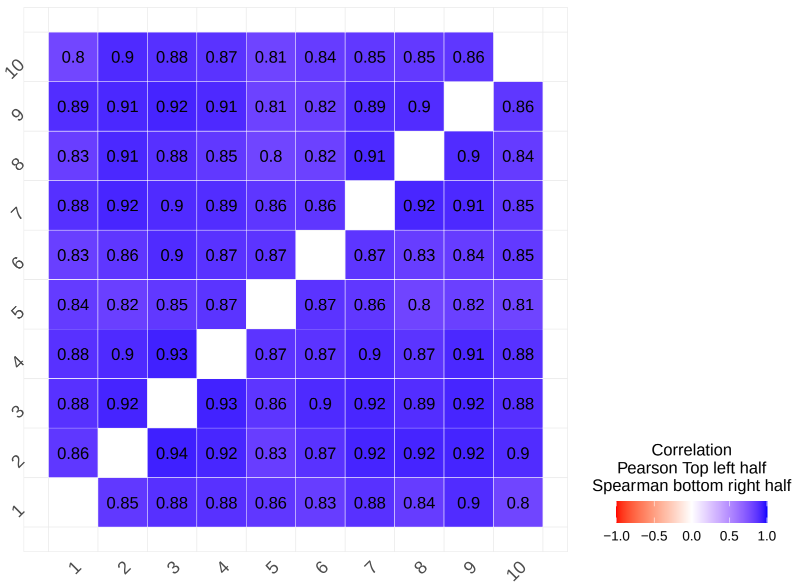
We first present tables and figures summarising the convergence path and fit of the model. The convergence plot of each seed with this specification can be seen in Figure J.15.1, while the correlation matrix between implied costs generated with each seed is displayed in Figure J.15.2. We summarise the final convergences achieved by each seed in Table J.15.1. The fraction of demands and costs that fall outside the feasible range is in Table J.15.2. Note that models that do not converge well may have more estimated probabilities outside the unit interval. Our analysis only uses the 5 trained models that have the highest R^2 however these poorly performing cases are not used. The variable importance of each of the model's input variables is summarised in Table J.15.3. We then present kernel densities of demand, cost, margin, o_1, o_2 and o_3 in Figure J.15.3. In Table J.15.4 (corresponding to Table 4) we provide the

Figure (J.15.1) Convergence Plot for Specification 15



Note: This chart plots the convergence of each trained model in terms of the R^2. We only plot positive R^2 for each which is why some seeds appear to start later on. In general, R^2 values start training at negative values.

Figure (J.15.2) Correlation Matrix Between Costs Implied by Each Seed for Specification 15



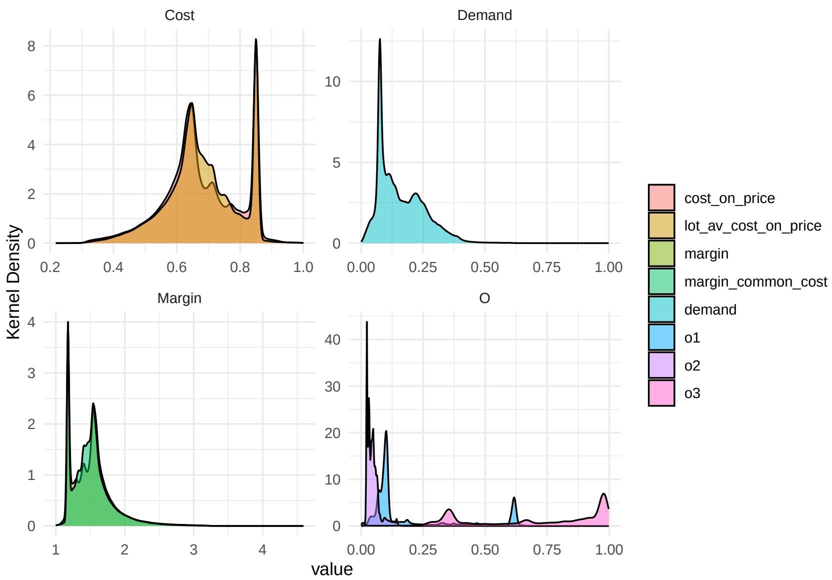
Note: This chart plots the correlation between $\frac{\text{Implied cost}}{\text{final price bid}}$ using the implied costs from each trained model. The top left half of the matrix are Pearson correlations while the bottom right half are Spearman correlations.

Table (J.15.3) Variable Importance Plot

Permuted Variable	Relative MSE	Relative Spearman Correlation
Price	15.9201	0.6365
Best Preceding Price	470.9973	0.727
Log Expected Price	1.0072	0.951
Num. Bidders after	1.0643	0.5014
Mean prices bidder	1.073	0.5882
Num. Participants	1.0313	0.7963
Prop. Inactive Bidders	1.062	0.6378
Price Reduction in Round 3	1.0015	0.9901
Round 2 std of bids	1.15	0.4293
Year tender published	1.0028	0.9837

Note: To construct this table we took the 5 trained models (of 10 in total) with above median performance on the test set. We recorded it's performance on the full (test + training) sample as a benchmark. We then permuted each of the input variables (30 times per model) and recorded their performance. We recorded the relative MSE and Spearman correlation for each permutation and each model. We then took the medians for each model. Finally we took the median across models to get one relative performance metric per permuted variable. A relative MSE above 1 indicates a deterioration in performance by permuting that variable while a relative spearman below 1 indicates the same.

Figure (J.15.3) Kernel Densities of Values Estimated with Specification 15



Note: This chart plots the kernel densities of demand, implied margin, cost and the o_1, o_2, o_3 parameters. For the purposes of this figure, we remove observations where $\frac{\text{Implied cost}}{\text{final price bid}}$ is outside the $[0, 1]$ range or demand is outside this same range. The amount of observations that are dropped in this way are small and represented in Table J.15.2.

regression coefficients of margin with different fixed effects for winners and losers.

Table (J.15.4) The Year-End Effect for Margins (with Varying Fixed Effects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Winners								
Last month	2.21***	2.13***	1.67***	1.61***	1.76***	1.72***	1.59***	1.67***
	(0.398)	(0.39)	(0.383)	(0.36)	(0.383)	(0.359)	(0.382)	(0.382)
R^2	0.00166	0.0318	0.301	0.142	0.303	0.164	0.404	0.406
N	136756	136751	136756	136756	136751	136751	136756	136751
Panel B: Losers								
Last month	3.49***	3.21***	2.71***	2.84***	2.8***	2.8***	2.56***	2.65***
	(0.519)	(0.506)	(0.458)	(0.452)	(0.458)	(0.451)	(0.436)	(0.436)
R^2	0.0025	0.0371	0.231	0.125	0.234	0.149	0.318	0.321
N	382559	382550	382559	382559	382550	382550	382559	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	No	Yes	No	No	Yes	Yes	No	Yes
Seller FE	No	No	Yes	No	Yes	No	Yes	Yes
Buyer FE	No	No	No	Yes	No	Yes	Yes	Yes

Note: The sample is composed of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is a margin common cost as defined in equation 12. Last month is a dummy variable that takes the value one if the tender was finalised in December or November and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and by seller. *p<0.1; **p<0.05; ***p<0.01.

The regressions with the interaction terms are in Table J.15.5 (which corresponds to Table 5).

Finally in Table J.15.5 we replicate the regressions without controls on different time intervals of data. Specifically, for all bids, winners and losers respectively we perform separate regressions: one on 2017-2019 (pre-Covid period) and one on 2020-2021 (where the bidding process was impacted by covid). We present our findings in Table J.15.6.

Table (J.15.5) The Year-End Effect for Margins (with Varying Controls)

	All bids			*Winners*			*Losers*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Last month	2.22***	1.54***	1.42***	1.67***	1.06***	1.29***	2.65***	1.88***	1.61***
	(0.383)	(0.36)	(0.427)	(0.382)	(0.368)	(0.482)	(0.436)	(0.406)	(0.488)
Last month × 4 bidders			0.665			0.209			0.897
			(0.607)			(0.73)			(0.652)
Last month × 5 bidders			1.29			0.264			1.64*
			(0.82)			(1.03)			(0.863)
Last month × 6 bidders			-1.2			-2.27*			-0.822
			(1.01)			(1.22)			(1.07)
Last month × 7+ bidders			-1.29			-1.98*			-1.33
			(0.983)			(1.18)			(1.05)
4 bidders		-0.0568	-0.161		-2.31***	-2.34***		0.217	0.0743
		(0.226)	(0.242)		(0.254)	(0.277)		(0.25)	(0.268)
5 bidders		-3.44***	-3.62***		-6.99***	-7.02***		-3.33***	-3.57***
		(0.298)	(0.307)		(0.348)	(0.371)		(0.318)	(0.329)
6 bidders		-7.71***	-7.55***		-11.4***	-11.1***		-7.76***	-7.66***
		(0.373)	(0.397)		(0.454)	(0.483)		(0.4)	(0.423)
7+ bidders		-13.3***	-13.1***		-14.2***	-13.9***		-14.2***	-14.1***
		(0.417)	(0.43)		(0.513)	(0.524)		(0.443)	(0.456)
R^2	0.29	0.301	0.301	0.406	0.418	0.418	0.321	0.334	0.334
N	519301	519301	519301	136751	136751	136751	382550	382550	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Columns 1, 4, and 7 correspond to specification 13. Columns 2, 5, and 8 also include dummies for each number of bidders in the auction. Columns 3, 6, and 9 further split the last month dummy by the number of bidders in the auction. Year FE stands for year fixed effects, buyer FE refers to government department fixed effects, seller FE indicates firm fixed effects, and sector FE represents sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

Table (J.15.6) The Year-End Effect for Margins (with Split by Year Interval)

	All 17-19	*All 20-21*	*Winners 17-19*	*Winners 20-21*	*Losers 17-19*	*Losers 20-21*
	(1)	(2)	(3)	(4)	(5)	(6)
Last month	2.19***	1.91***	2.0***	1.22**	2.5***	2.26***
	(0.493)	(0.5)	(0.587)	(0.518)	(0.57)	(0.571)
R^2	0.373	0.34	0.501	0.49	0.414	0.374
N	208764	310537	56309	80442	152455	230095
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation 12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

J.16 Specification 16

This specification mirrors the benchmark specification, with the exception that the neural network is trained including the bidders who participate last in the auction. The presence of these bidders is only used in the training of the neural network. Since their demand functions are likely to be stepped (as undercutting the previous bidder by an epsilon is likely to discontinuously increase the probability of winning), we exclude the inferred costs from these bidders in the regressions on margin. Although this approach allows for a broader training dataset, it is not our preferred specification. Training on cases that we ultimately are not using could detract from the model performance on cases that are important for our paper. Furthermore, accuracy in these cases can influence the early stopping logic of the model. These results do show an R^2 around 37% however which suggests that the reason other R^2s are lower is due to the removal of the easier cases from the training and test datasets.

Table (J.16.1) Final convergence

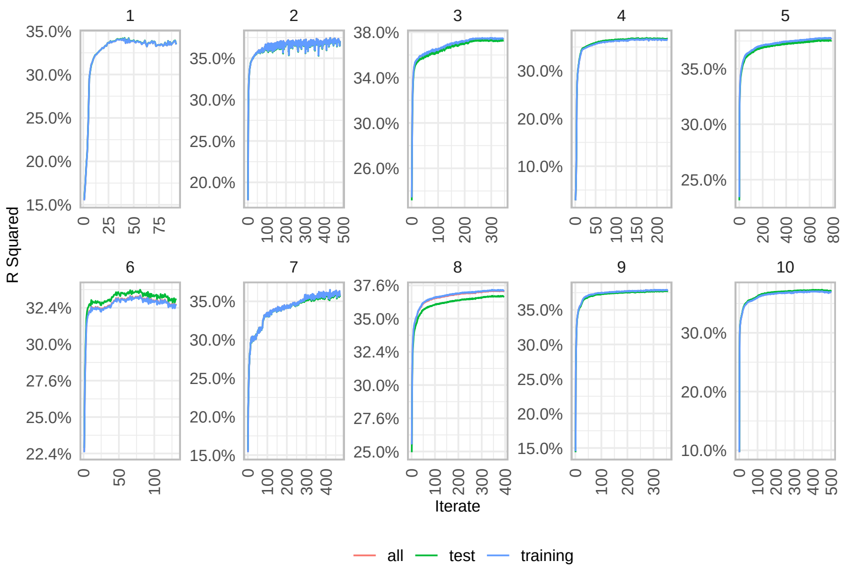
		R^2			Pearson Corr			MAE		
Seed	Num. Iterates	All	Test	Training	All	Test	Training	All	Test	Training
1	92	0.3346	0.3348	0.3346	0.5997	0.6002	0.5996	0.3231	0.3235	0.3231
2	483	0.371	0.3708	0.3711	0.6111	0.611	0.6111	0.2876	0.2879	0.2876
3	347	0.3744	0.3728	0.3746	0.6122	0.6108	0.6124	0.2973	0.298	0.2972
4	226	0.3653	0.3671	0.3651	0.6059	0.6075	0.6058	0.3036	0.3035	0.3036
5	785	0.3774	0.3755	0.3776	0.6154	0.6137	0.6155	0.292	0.2924	0.292
6	131	0.3277	0.3313	0.3273	0.5873	0.5906	0.587	0.3241	0.3231	0.3242
7	472	0.3568	0.3558	0.357	0.6085	0.6074	0.6086	0.3084	0.3086	0.3084
8	394	0.3709	0.3669	0.3714	0.6109	0.6073	0.6112	0.3006	0.301	0.3005
9	354	0.379	0.3777	0.3792	0.6167	0.6155	0.6168	0.297	0.2969	0.297
10	501	0.3689	0.371	0.3687	0.6082	0.61	0.608	0.3014	0.3006	0.3015

Note: This table shows the R^2, Pearson correlation and Mean Absolute Error (MAE) of the prediction at the point when training stops. Due to the early stopping logic, training stops after 50 iterates in which performance on the loss function on the test set has not improved.

We first present tables and figures summarising the convergence path and fit of the model. The convergence plot of each seed with this specification can be seen in Figure J.16.1, while the correlation matrix between implied costs generated with each seed is displayed in Figure J.16.2. We summarise the final convergences achieved by each seed in Table J.16.1. The fraction of demands and costs that fall outside the feasible range is in Table J.16.2. Note that models that do not converge well may have more estimated probabilities outside the unit interval. Our analysis only uses the 5 trained models that have the highest R^2 however these poorly performing cases are not used. The variable importance of each of the model's input variables is summarised in Table J.16.3. We then present kernel densities of demand, cost, margin, o_1, o_2 and o_3 in Figure J.16.3. In Table J.16.4 (corresponding to Table 4) we provide the regression coefficients of margin with different fixed effects for winners and losers.

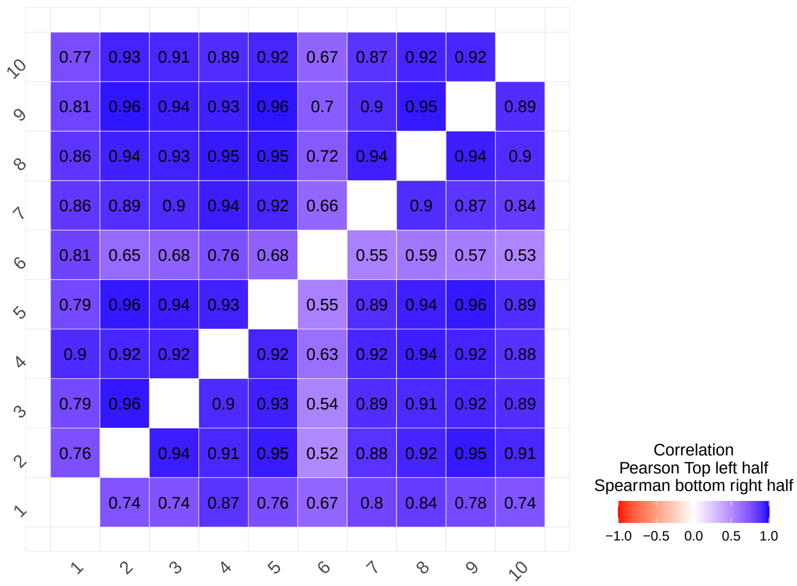
The regressions with the interaction terms are in Table J.16.5 (which corresponds to Table 5).

Figure (J.16.1) Convergence Plot for Specification 16



Note: This chart plots the convergence of each trained model in terms of the R^2. We only plot positive R^2 for each which is why some seeds appear to start later on. In general, R^2 values start training at negative values.

Figure (J.16.2) Correlation Matrix Between Costs Implied by Each Seed for Specification 16



Note: This chart plots the correlation between $\frac{\text{Implied cost}}{\text{final price bid}}$ using the implied costs from each trained model. The top left half of the matrix are Pearson correlations while the bottom right half are Spearman correlations.

Table (J.16.2) Proportion of invalid values for each trained model.

Seed	Demand > 1	Demand < 0	Cost > Bid	Cost < 0
Combined	0.0001%	0.0001%	0.0022%	0
1	0.0006%	0.0006%	0%	0
2	0%	0%	0.1647%	0
3	0.0002%	0.0002%	0.0039%	0
4	0.0005%	0.0005%	0%	0
5	0.0001%	0.0001%	0%	0
6	0%	0%	0%	0
7	0%	0%	0%	0
8	0%	0%	0%	0
9	0.0014%	0.0014%	0.0036%	0
10	0.0018%	0.0018%	0%	0

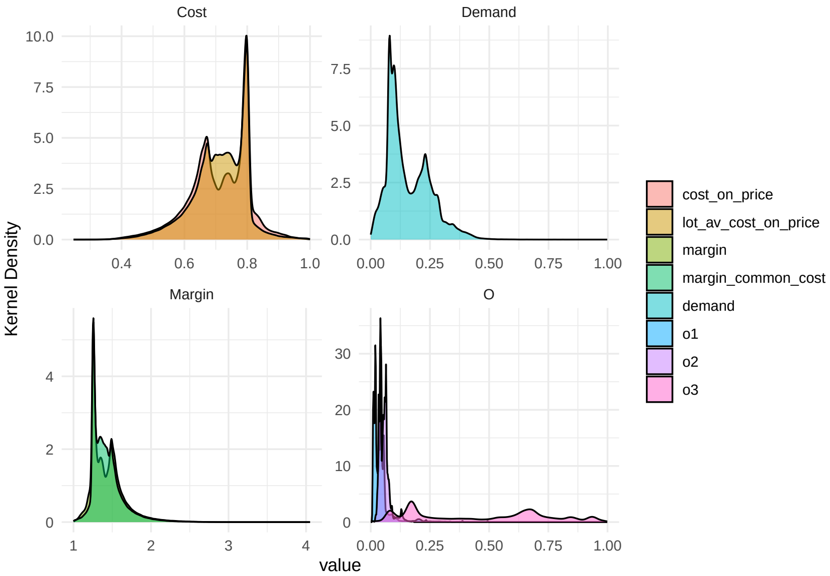
Note: This table shows fraction of demands that are outside the unit interval and the fraction of costs that are above the final bid cost or below zero. Each figure in this table is a percentage of the non-missing values in the dataset. Figures are expressed to 4 decimal places.

Table (J.16.3) Variable Importance Plot

Permuted Variable	Relative MSE	Relative Spearman Correlation
Price	3125.8442	0.8581
Best Preceding Price	7456177.1766	0.7717
Log Expected Price	1.0075	0.9938
Num. Bidders after	1.9869	0.1026
Mean prices bidder	1.0598	0.9468
Num. Participants	1.0195	0.9827
Prop. Inactive Bidders	1.0445	0.9592
Price Reduction in Round 3	1.0046	0.9966
Round 2 std of bids	1.1789	0.8486
Year tender published	1.0026	0.9979

Note: To construct this table we took the 5 trained models (of 10 in total) with above median performance on the test set. We recorded it's performance on the full (test + training) sample as a benchmark. We then permuted each of the input variables (30 times per model) and recorded their performance. We recorded the relative MSE and Spearman correlation for each permutation and each model. We then took the medians for each model. Finally we took the median across models to get one relative performance metric per permuted variable. A relative MSE above 1 indicates a deterioration in performance by permuting that variable while a relative spearman below 1 indicates the same.

Figure (J.16.3)   Kernel Densities of Values Estimated with Specification 16



Note: This chart plots the kernel densities of demand, implied margin, cost and the o_1, o_2, o_3 parameters. For the purposes of this figure, we remove observations where $\frac{\text{Implied cost}}{\text{final price bid}}$ is outside the $[0, 1]$ range or demand is outside this same range. The amount of observations that are dropped in this way are small and represented in Table J.16.2.

Finally in Table J.16.5 we replicate the regressions without controls on different time intervals of data. Specifically, for all bids, winners and losers respectively we perform separate regressions: one on 2017-2019 (pre-Covid period) and one on 2020-2021 (where the bidding process was impacted by covid). We present our findings in Table J.16.6.

Table (J.16.4) The Year-End Effect for Margins (with Varying Fixed Effects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Winners								
Last month	2.11***	1.95***	1.67***	1.64***	1.74***	1.66***	1.65***	1.71***
	(0.302)	(0.295)	(0.301)	(0.285)	(0.302)	(0.285)	(0.309)	(0.309)
R^2	0.0011	0.0235	0.288	0.14	0.289	0.155	0.393	0.395
N	136756	136751	136756	136756	136751	136751	136756	136751
Panel B: Losers								
Last month	3.25***	2.83***	2.49***	2.71***	2.54***	2.56***	2.44***	2.49***
	(0.391)	(0.378)	(0.349)	(0.347)	(0.349)	(0.346)	(0.34)	(0.34)
R^2	0.00233	0.0265	0.218	0.118	0.221	0.133	0.305	0.307
N	382559	382550	382559	382559	382550	382550	382559	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	No	Yes	No	No	Yes	Yes	No	Yes
Seller FE	No	No	Yes	No	Yes	No	Yes	Yes
Buyer FE	No	No	No	Yes	No	Yes	Yes	Yes

Note: The sample is composed of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is a margin common cost as defined in equation 12. Last month is a dummy variable that takes the value one if the tender was finalised in December or November and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and by seller. *p<0.1; **p<0.05; ***p<0.01.

Table (J.16.5) The Year-End Effect for Margins (with Varying Controls)

	All bids			*Winners*			*Losers*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Last month	2.16***	1.33***	1.39***	1.71***	0.938***	1.25***	2.49***	1.59***	1.59***
	(0.301)	(0.279)	(0.369)	(0.309)	(0.294)	(0.413)	(0.34)	(0.314)	(0.419)
Last month × 4 bidders			0.21			-0.00869			0.328
			(0.529)			(0.626)			(0.57)
Last month × 5 bidders			0.504			-0.42			0.691
			(0.681)			(0.849)			(0.721)
Last month × 6 bidders			-1.0			-2.0**			-0.766
			(0.832)			(0.975)			(0.884)
Last month × 7+ bidders			-0.934			-1.4			-1.07
			(0.746)			(0.92)			(0.799)
4 bidders		-3.64***	-3.67***		-5.31***	-5.31***		-3.51***	-3.56***
		(0.202)	(0.217)		(0.222)	(0.242)		(0.223)	(0.24)
5 bidders		-7.94***	-8.01***		-10.8***	-10.7***		-7.89***	-7.99***
		(0.253)	(0.261)		(0.298)	(0.318)		(0.268)	(0.279)
6 bidders		-11.8***	-11.6***		-14.8***	-14.5***		-11.9***	-11.8***
		(0.322)	(0.342)		(0.386)	(0.412)		(0.345)	(0.364)
7+ bidders		-16.5***	-16.4***		-18.2***	-18.0***		-17.1***	-17.0***
		(0.326)	(0.339)		(0.392)	(0.404)		(0.35)	(0.363)
R^2	0.274	0.298	0.298	0.395	0.423	0.423	0.307	0.33	0.33
N	519301	519301	519301	136751	136751	136751	382550	382550	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Columns 1, 4, and 7 correspond to specification 13. Columns 2, 5, and 8 also include dummies for each number of bidders in the auction. Columns 3, 6, and 9 further split the last month dummy by the number of bidders in the auction. Year FE stands for year fixed effects, buyer FE refers to government department fixed effects, seller FE indicates firm fixed effects, and sector FE represents sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

Table (J.16.6) The Year-End Effect for Margins (with Split by Year Interval)

	All 17-19	*All 20-21*	*Winners 17-19*	*Winners 20-21*	*Losers 17-19*	*Losers 20-21*
	(1)	(2)	(3)	(4)	(5)	(6)
Last month	2.33***	1.88***	2.02***	1.39***	2.59***	2.15***
	(0.406)	(0.401)	(0.5)	(0.427)	(0.467)	(0.455)
R^2	0.361	0.324	0.49	0.482	0.403	0.358
N	208764	310537	56309	80442	152455	230095
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation 12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

J.17 Specification 17

This makes a number of deviations away from the benchmark specification at the same time. Specifically alternating layers of Tanh and Swish activation functions are used, there is 4% dropout probability per node, the loss function is MSE, the cost floor is 0.2 and there are 5 hidden layers.

Table (J.17.1) Final convergence

		R^2			Pearson Corr			MAE		
Seed	Num. Iterates	All	Test	Training	All	Test	Training	All	Test	Training
1	472	0.0477	0.0493	0.0475	0.2193	0.2228	0.2189	0.2652	0.2653	0.2652
2	791	0.0589	0.0591	0.0589	0.2435	0.2435	0.2435	0.2624	0.2635	0.2623
3	1538	0.0673	0.0676	0.0672	0.26	0.2609	0.2599	0.2548	0.2552	0.2548
4	845	0.066	0.0643	0.0662	0.2589	0.2551	0.2593	0.2538	0.2542	0.2538
5	1891	0.0701	0.0701	0.0701	0.2658	0.2656	0.2658	0.2554	0.2553	0.2554
6	249	0.0043	0.0022	0.0046	0.1202	0.1162	0.1206	0.2813	0.2805	0.2814
7	354	0.0357	0.0363	0.0356	0.1991	0.2006	0.199	0.2752	0.2745	0.2752
8	2068	0.0657	0.0646	0.0659	0.2578	0.2555	0.258	0.2532	0.2542	0.2531
9	1633	0.0702	0.0712	0.0701	0.2662	0.268	0.266	0.2552	0.2543	0.2553
10	1462	0.0689	0.0661	0.0692	0.2632	0.2574	0.2638	0.2556	0.2551	0.2556

Note: This table shows the R^2, Pearson correlation and Mean Absolute Error (MAE) of the prediction at the point when training stops. Due to the early stopping logic, training stops after 50 iterates in which performance on the loss function on the test set has not improved.

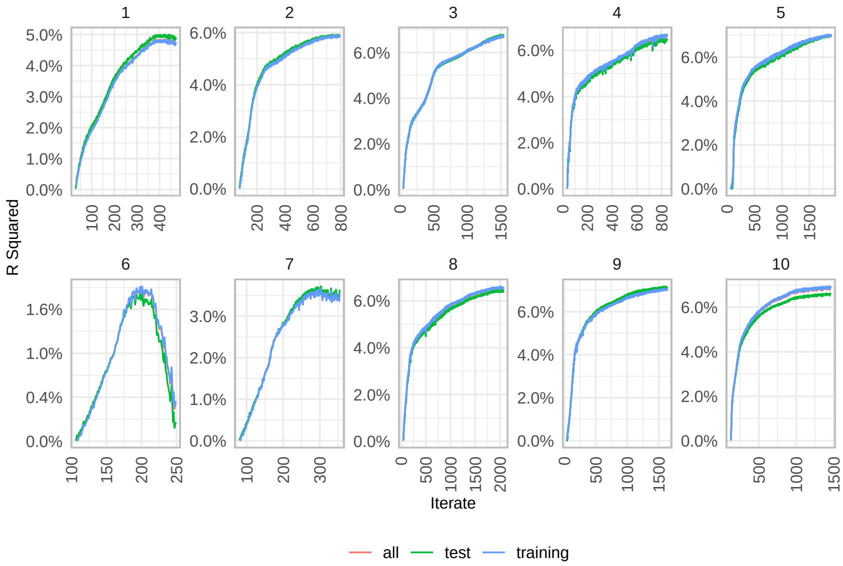
Table (J.17.2) Proportion of invalid values for each trained model.

Seed	Demand > 1	Demand < 0	Cost > Bid	Cost < 0
Combined	0%	0%	0.0004%	0
1	0%	0%	0%	0
2	0.0001%	0.0001%	0.1525%	0
3	0%	0%	0.014%	0
4	0.0006%	0.0006%	0.7356%	0
5	0.0004%	0.0004%	0.0045%	0
6	0%	0%	0.0868%	0
7	0.0001%	0.0001%	0.4431%	0
8	0.0002%	0.0002%	0.0402%	0
9	0.0002%	0.0002%	0.8046%	0
10	0.0004%	0.0004%	0%	0

Note: This table shows fraction of demands that are outside the unit interval and the fraction of costs that are above the final bid cost or below zero. Each figure in this table is a percentage of the non-missing values in the dataset. Figures are expressed to 4 decimal places.

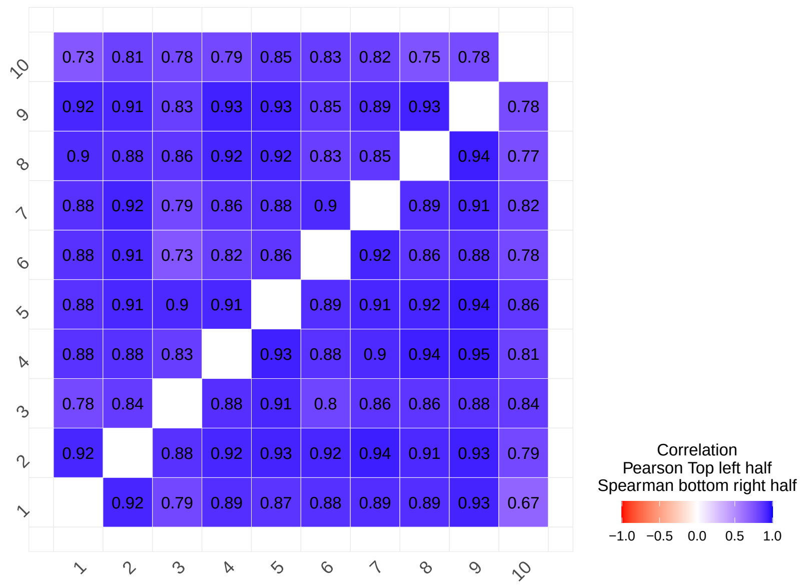
We first present tables and figures summarising the convergence path and fit of the model. The convergence plot of each seed with this specification can be seen in Figure J.17.1, while the correlation matrix between implied costs generated with each seed is displayed in Figure J.17.2. We summarise the final convergences achieved by each seed in Table J.17.1. The fraction of demands and costs that fall outside the feasible range is in Table J.17.2. Note that models that do not converge well may have more estimated probabilities outside the unit interval. Our analysis only uses the 5 trained models that have

Figure (J.17.1) Convergence Plot for Specification 17



Note: This chart plots the convergence of each trained model in terms of the R^2. We only plot positive R^2 for each which is why some seeds appear to start later on. In general, R^2 values start training at negative values.

Figure (J.17.2) Correlation Matrix Between Costs Implied by Each Seed for Specification 17



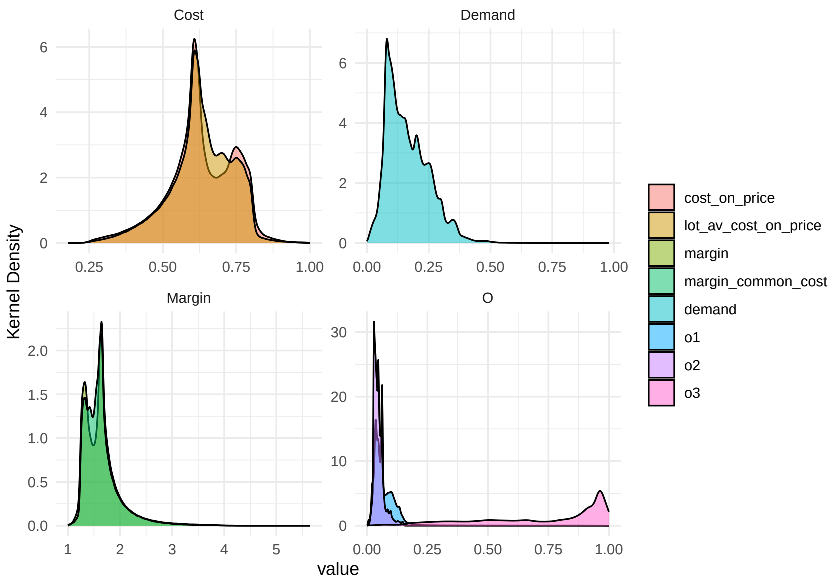
Note: This chart plots the correlation between $\frac{\text{Implied cost}}{\text{final price bid}}$ using the implied costs from each trained model. The top left half of the matrix are Pearson correlations while the bottom right half are Spearman correlations.

Table (J.17.3) Variable Importance Plot

Permuted Variable	Relative MSE	Relative Spearman Correlation
Price	4.3105	0.6972
Best Preceding Price	3.4462	0.7366
Log Expected Price	1.006	0.9569
Num. Bidders after	1.0581	0.5389
Mean prices bidder	1.0665	0.581
Num. Participants	1.0371	0.7699
Prop. Inactive Bidders	1.0484	0.6418
Price Reduction in Round 3	1.0012	0.9927
Round 2 std of bids	1.1316	0.4606
Year tender published	1.0025	0.9852

Note: To construct this table we took the 5 trained models (of 10 in total) with above median performance on the test set. We recorded it's performance on the full (test + training) sample as a benchmark. We then permuted each of the input variables (30 times per model) and recorded their performance. We recorded the relative MSE and Spearman correlation for each permutation and each model. We then took the medians for each model. Finally we took the median across models to get one relative performance metric per permuted variable. A relative MSE above 1 indicates a deterioration in performance by permuting that variable while a relative spearman below 1 indicates the same.

Figure (J.17.3) Kernel Densities of Values Estimated with Specification 17



Note: This chart plots the kernel densities of demand, implied margin, cost and the o_1, o_2, o_3 parameters. For the purposes of this figure, we remove observations where $\frac{\text{Implied cost}}{\text{final price bid}}$ is outside the $[0, 1]$ range or demand is outside this same range. The amount of observations that are dropped in this way are small and represented in Table J.17.2.

the highest R^2 however these poorly performing cases are not used. The variable importance of each of the model's input variables is summarised in Table J.17.3. We then present kernel densities of demand, cost, margin, o_1, o_2 and o_3 in Figure J.17.3. In Table J.17.4 (corresponding to Table 4) we provide the regression coefficients of margin with different fixed effects for winners and losers.

Table (J.17.4) The Year-End Effect for Margins (with Varying Fixed Effects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Winners								
Last month	2.88***	2.86***	2.25***	2.2***	2.37***	2.36***	2.13***	2.24***
	(0.518)	(0.504)	(0.491)	(0.461)	(0.49)	(0.459)	(0.487)	(0.487)
R^2	0.00152	0.033	0.301	0.141	0.303	0.164	0.404	0.406
N	136756	136751	136756	136756	136751	136751	136756	136751
Panel B: Losers								
Last month	4.31***	4.09***	3.51***	3.63***	3.63***	3.61***	3.3***	3.42***
	(0.681)	(0.66)	(0.599)	(0.586)	(0.599)	(0.581)	(0.564)	(0.563)
R^2	0.00196	0.0389	0.232	0.128	0.235	0.154	0.321	0.325
N	382559	382550	382559	382559	382550	382550	382559	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	No	Yes	No	No	Yes	Yes	No	Yes
Seller FE	No	No	Yes	No	Yes	No	Yes	Yes
Buyer FE	No	No	No	Yes	No	Yes	Yes	Yes

Note: The sample is composed of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is a margin common cost as defined in equation 12. Last month is a dummy variable that takes the value one if the tender was finalised in December or November and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and by seller. *p<0.1; **p<0.05; ***p<0.01.

The regressions with the interaction terms are in Table J.17.5 (which corresponds to Table 5).

Finally in Table J.17.5 we replicate the regressions without controls on different time intervals of data. Specifically, for all bids, winners and losers respectively we perform separate regressions: one on 2017-2019 (pre-Covid period) and one on 2020-2021 (where the bidding process was impacted by covid). We present our findings in Table J.17.6.

Table (J.17.5) The Year-End Effect for Margins (with Varying Controls)

	All bids			*Winners*			*Losers*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Last month	2.92***	2.11***	2.03***	2.24***	1.5***	1.87***	3.42***	2.51***	2.25***
	(0.494)	(0.468)	(0.56)	(0.487)	(0.471)	(0.619)	(0.563)	(0.529)	(0.642)
Last month × 4 bidders			0.68			0.0242			0.99
			(0.781)			(0.927)			(0.838)
Last month × 5 bidders			1.41			0.367			1.78
			(1.05)			(1.32)			(1.1)
Last month × 6 bidders			-1.38			-2.59*			-0.948
			(1.3)			(1.55)			(1.38)
Last month × 7+ bidders			-1.75			-2.89*			-1.67
			(1.26)			(1.49)			(1.35)
4 bidders		-0.794***	-0.901***		-3.27***	-3.27***		-0.504	-0.661**
		(0.285)	(0.306)		(0.319)	(0.348)		(0.313)	(0.336)
5 bidders		-5.07***	-5.27***		-9.09***	-9.14***		-4.96***	-5.22***
		(0.372)	(0.385)		(0.431)	(0.46)		(0.396)	(0.412)
6 bidders		-9.49***	-9.31***		-14.2***	-13.8***		-9.43***	-9.31***
		(0.472)	(0.505)		(0.558)	(0.599)		(0.506)	(0.537)
7+ bidders		-15.9***	-15.7***		-16.8***	-16.5***		-17.0***	-16.8***
		(0.519)	(0.535)		(0.616)	(0.629)		(0.552)	(0.568)
R^2	0.294	0.304	0.304	0.406	0.417	0.417	0.325	0.335	0.335
N	519301	519301	519301	136751	136751	136751	382550	382550	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Columns 1, 4, and 7 correspond to specification 13. Columns 2, 5, and 8 also include dummies for each number of bidders in the auction. Columns 3, 6, and 9 further split the last month dummy by the number of bidders in the auction. Year FE stands for year fixed effects, buyer FE refers to government department fixed effects, seller FE indicates firm fixed effects, and sector FE represents sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

Table (J.17.6) The Year-End Effect for Margins (with Split by Year Interval)

	All 17-19	*All 20-21*	*Winners 17-19*	*Winners 20-21*	*Losers 17-19*	*Losers 20-21*
	(1)	(2)	(3)	(4)	(5)	(6)
Last month	2.84***	2.6***	2.69***	1.64**	3.12***	3.06***
	(0.632)	(0.639)	(0.753)	(0.662)	(0.726)	(0.726)
R^2	0.377	0.346	0.5	0.49	0.417	0.379
N	208764	310537	56309	80442	152455	230095
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation 12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

J.18 Specification 18

This is similar to the benchmark specification but does not have three of the weakest inputs (according to variable importance measures). The excluded inputs are year tender publishing date; reduction in price so far in round 3; and log of the expected price.

Table (J.18.1) Final convergence

		R^2			Pearson Corr			MAE		
Seed	Num. Iterates	All	Test	Training	All	Test	Training	All	Test	Training
1	87	0.0348	0.0365	0.0346	0.1869	0.1911	0.1864	0.2619	0.262	0.2619
2	920	0.059	0.0604	0.0588	0.2433	0.246	0.243	0.2601	0.2611	0.26
3	708	0.0532	0.0532	0.0532	0.2341	0.2346	0.234	0.2512	0.2514	0.2511
4	303	0.0545	0.0536	0.0545	0.2345	0.2326	0.2347	0.2542	0.2544	0.2542
5	201	0.0454	0.044	0.0456	0.2156	0.2121	0.216	0.253	0.2531	0.253
6	694	0.0538	0.0538	0.0538	0.2329	0.2326	0.2329	0.2569	0.256	0.257
7	247	0.0354	0.0388	0.035	0.1905	0.1991	0.1895	0.2561	0.2551	0.2562
8	313	0.0532	0.0525	0.0533	0.232	0.2306	0.2322	0.2528	0.2538	0.2527
9	594	0.0491	0.0494	0.049	0.2271	0.2272	0.2271	0.2503	0.2495	0.2504
10	172	0.0478	0.0461	0.048	0.222	0.2174	0.2225	0.252	0.2512	0.2521

Note: This table shows the R^2, Pearson correlation and Mean Absolute Error (MAE) of the prediction at the point when training stops. Due to the early stopping logic, training stops after 50 iterates in which performance on the loss function on the test set has not improved.

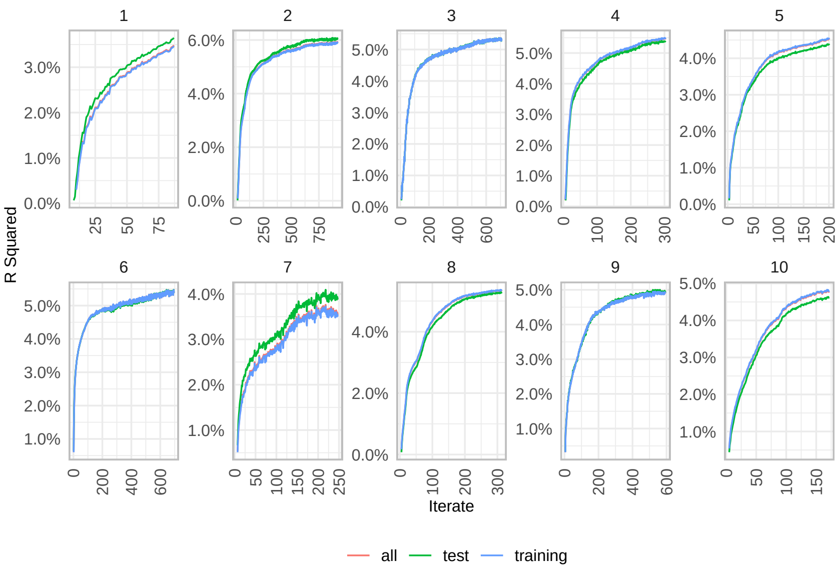
Table (J.18.2) Proportion of invalid values for each trained model.

Seed	Demand > 1	Demand < 0	Cost > Bid	Cost < 0
Combined	0%	0%	0%	0%
1	0.0006%	0.0006%	0%	0%
2	0.0034%	0.0034%	0.0018%	0%
3	0%	0%	0.0269%	0%
4	0%	0%	0.2111%	0%
5	0%	0%	0%	0%
6	0%	0%	0.0022%	0%
7	0.0002%	0.0002%	0.0909%	0%
8	0%	0%	0%	0%
9	0.0005%	0.0005%	0%	0%
10	0.0001%	0.0001%	0%	0%

Note: This table shows fraction of demands that are outside the unit interval and the fraction of costs that are above the final bid cost or below zero. Each figure in this table is a percentage of the non-missing values in the dataset. Figures are expressed to 4 decimal places.

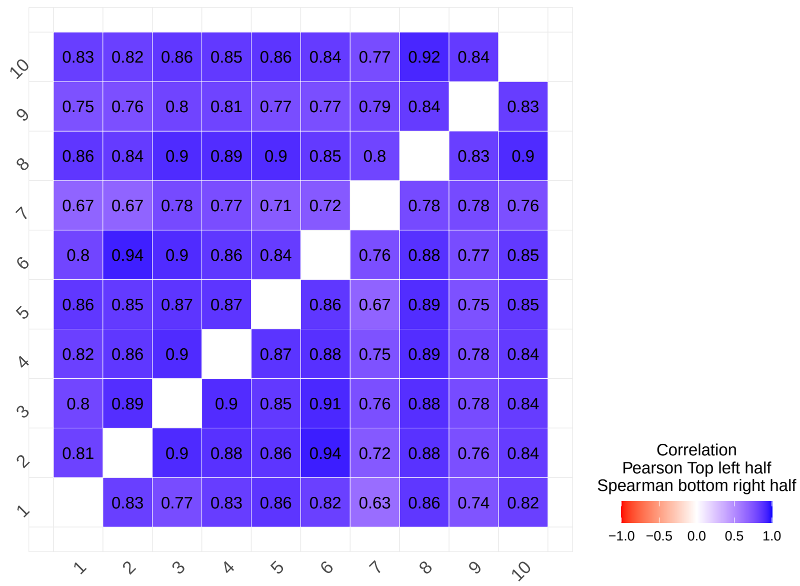
We first present tables and figures summarising the convergence path and fit of the model. The convergence plot of each seed with this specification can be seen in Figure J.18.1, while the correlation matrix between implied costs generated with each seed is displayed in Figure J.18.2. We summarise the final convergences achieved by each seed in Table J.18.1. The fraction of demands and costs that fall outside the feasible range is in Table J.18.2. Note that models that do not converge well may have more estimated probabilities outside the unit interval. Our analysis only uses the 5 trained models that have the highest R^2 however these poorly performing cases are not used. The variable importance of each of

Figure (J.18.1) Convergence Plot for Specification 18



Note: This chart plots the convergence of each trained model in terms of the R^2. We only plot positive R^2 for each which is why some seeds appear to start later on. In general, R^2 values start training at negative values.

Figure (J.18.2) Correlation Matrix Between Costs Implied by Each Seed for Specification 18



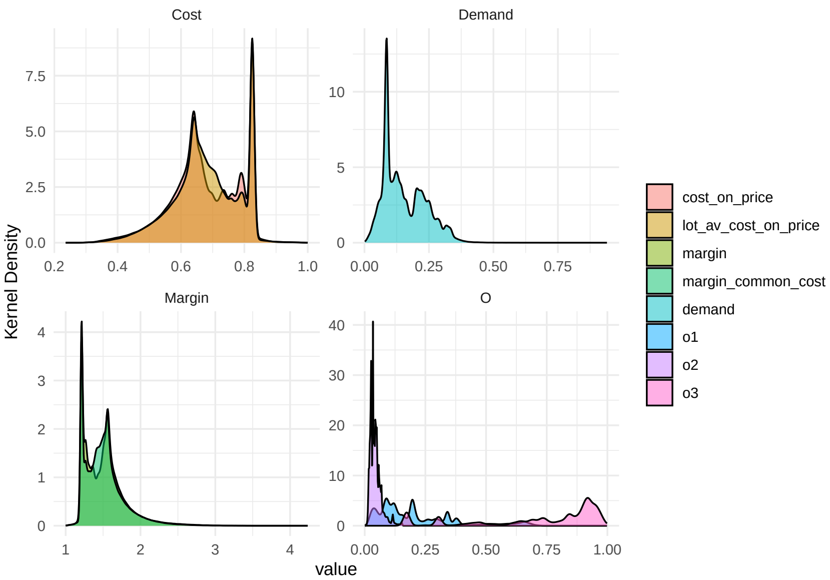
Note: This chart plots the correlation between $\frac{\text{Implied cost}}{\text{final price bid}}$ using the implied costs from each trained model. The top left half of the matrix are Pearson correlations while the bottom right half are Spearman correlations.

Table (J.18.3) Variable Importance Plot

Permuted Variable	Relative MSE	Relative Spearman Correlation
Price	15.0514	0.6023
Best Preceding Price	333.8106	0.6862
Num. Bidders after	1.047	0.5412
Mean prices bidder	1.0478	0.6352
Num. Participants	1.0132	0.8997
Prop. Inactive Bidders	1.0425	0.6622
Round 2 std of bids	1.1279	0.3613

Note: To construct this table we took the 5 trained models (of 10 in total) with above median performance on the test set. We recorded it's performance on the full (test + training) sample as a benchmark. We then permuted each of the input variables (30 times per model) and recorded their performance. We recorded the relative MSE and Spearman correlation for each permutation and each model. We then took the medians for each model. Finally we took the median across models to get one relative performance metric per permuted variable. A relative MSE above 1 indicates a deterioration in performance by permuting that variable while a relative spearman below 1 indicates the same.

Figure (J.18.3) Kernel Densities of Values Estimated with Specification 18



Note: This chart plots the kernel densities of demand, implied margin, cost and the o_1, o_2, o_3 parameters. For the purposes of this figure, we remove observations where $\frac{\text{Implied cost}}{\text{final price bid}}$ is outside the $[0, 1]$ range or demand is outside this same range. The amount of observations that are dropped in this way are small and represented in Table J.18.2.

the model's input variables is summarised in Table J.18.3. We then present kernel densities of demand, cost, margin, o_1, o_2 and o_3 in Figure J.18.3. In Table J.18.4 (corresponding to Table 4) we provide the regression coefficients of margin with different fixed effects for winners and losers.

Table (J.18.4) The Year-End Effect for Margins (with Varying Fixed Effects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Winners								
Last month	2.25***	2.14***	1.7***	1.63***	1.78***	1.73***	1.63***	1.7***
	(0.364)	(0.357)	(0.352)	(0.328)	(0.352)	(0.33)	(0.351)	(0.351)
R^2	0.00253	0.0305	0.298	0.145	0.3	0.165	0.403	0.405
N	136756	136751	136756	136756	136751	136751	136756	136751
Panel B: Losers								
Last month	3.48***	3.15***	2.67***	2.77***	2.75***	2.73***	2.55***	2.62***
	(0.476)	(0.465)	(0.424)	(0.408)	(0.424)	(0.409)	(0.4)	(0.399)
R^2	0.00557	0.0389	0.233	0.128	0.236	0.15	0.32	0.323
N	382559	382550	382559	382559	382550	382550	382559	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	No	Yes	No	No	Yes	Yes	No	Yes
Seller FE	No	No	Yes	No	Yes	No	Yes	Yes
Buyer FE	No	No	No	Yes	No	Yes	Yes	Yes

Note: The sample is composed of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is a margin common cost as defined in equation 12. Last month is a dummy variable that takes the value one if the tender was finalised in December or November and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and by seller. *p<0.1; **p<0.05; ***p<0.01.

The regressions with the interaction terms are in Table J.18.5 (which corresponds to Table 5).

Finally in Table J.18.5 we replicate the regressions without controls on different time intervals of data. Specifically, for all bids, winners and losers respectively we perform separate regressions: one on 2017-2019 (pre-Covid period) and one on 2020-2021 (where the bidding process was impacted by covid). We present our findings in Table J.18.6.

Table (J.18.5) The Year-End Effect for Margins (with Varying Controls)

	All bids (1)	(2)	(3)	*Winners* (4)	(5)	(6)	*Losers* (7)	(8)	(9)
Last month	2.22***	1.61***	1.51***	1.7***	1.07***	1.4***	2.62***	1.94***	1.72***
	(0.351)	(0.332)	(0.42)	(0.351)	(0.337)	(0.459)	(0.399)	(0.375)	(0.481)
Last month × 4 bidders			0.439			0.00701			0.629
			(0.58)			(0.688)			(0.625)
Last month × 5 bidders			0.881			-0.333			1.16
			(0.768)			(0.936)			(0.814)
Last month × 6 bidders			-0.868			-2.11*			-0.537
			(0.915)			(1.09)			(0.975)
Last month × 7+ bidders			-0.756			-1.71*			-0.742
			(0.82)			(0.98)			(0.887)
4 bidders		-0.281	-0.35		-2.52***	-2.52***		0.00878	-0.0915
		(0.217)	(0.232)		(0.239)	(0.26)		(0.24)	(0.257)
5 bidders		-2.95***	-3.07***		-6.6***	-6.55***		-2.75***	-2.92***
		(0.276)	(0.284)		(0.32)	(0.34)		(0.294)	(0.305)
6 bidders		-6.8***	-6.69***		-10.9***	-10.6***		-6.7***	-6.64***
		(0.342)	(0.363)		(0.408)	(0.436)		(0.366)	(0.387)
7+ bidders		-12.3***	-12.2***		-15.3***	-15.1***		-12.7***	-12.6***
		(0.334)	(0.348)		(0.407)	(0.419)		(0.361)	(0.376)
R^2	0.29	0.301	0.301	0.405	0.42	0.42	0.323	0.334	0.334
N	519301	519301	519301	136751	136751	136751	382550	382550	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Columns 1, 4, and 7 correspond to specification 13. Columns 2, 5, and 8 also include dummies for each number of bidders in the auction. Columns 3, 6, and 9 further split the last month dummy by the number of bidders in the auction. Year FE stands for year fixed effects, buyer FE refers to government department fixed effects, seller FE indicates firm fixed effects, and sector FE represents sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

Table (J.18.6) The Year-End Effect for Margins (with Split by Year Interval)

	All 17-19 (1)	*All 20-21* (2)	*Winners 17-19* (3)	*Winners 20-21* (4)	*Losers 17-19* (5)	*Losers 20-21* (6)
Last month	2.28***	1.87***	1.94***	1.25***	2.57***	2.19***
	(0.444)	(0.468)	(0.545)	(0.481)	(0.516)	(0.534)
R^2	0.373	0.338	0.5	0.489	0.414	0.372
N	208764	310537	56309	80442	152455	230095
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation 12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

J.19 Specification 19

This is similar to the benchmark specification but has two additional inputs to the neural network: the size of the firm (with the categories "not a business", "micro", "small", "medium", and "large" encoded as -2, -1, 0, 1, 2) and the number of days since the firm was registered. The logic behind these fields is that larger and longer-established firms might be preferred by buyers and hence face a flatter demand curve. These were not included in the benchmark specification as we have limited data on these fields and need to remove some observations to accommodate them.

Table (J.19.1) Final convergence

		R^2			Pearson Corr			MAE		
Seed	Num. Iterates	All	Test	Training	All	Test	Training	All	Test	Training
1	447	0.0593	0.0605	0.0592	0.245	0.2486	0.2446	0.2563	0.2579	0.2562
2	835	0.0713	0.0698	0.0714	0.2672	0.2645	0.2675	0.2571	0.2573	0.257
3	1080	0.0663	0.0679	0.0661	0.2578	0.2614	0.2574	0.2575	0.2593	0.2574
4	1323	0.0723	0.0749	0.072	0.2706	0.2764	0.2699	0.2549	0.2556	0.2548
5	1036	0.072	0.0712	0.072	0.2688	0.2672	0.2689	0.2568	0.2564	0.2569
6	155	0.0413	0.0415	0.0413	0.2049	0.2057	0.2048	0.2601	0.2607	0.26
7	625	0.058	0.0583	0.0579	0.241	0.2414	0.241	0.2631	0.2644	0.263
8	965	0.0681	0.0663	0.0683	0.2615	0.2579	0.2619	0.2564	0.2565	0.2564
9	370	0.0584	0.0585	0.0584	0.2426	0.2432	0.2426	0.2566	0.2568	0.2565
10	417	0.0443	0.0435	0.0444	0.2113	0.2094	0.2115	0.2609	0.2615	0.2608

Note: This table shows the R^2, Pearson correlation and Mean Absolute Error (MAE) of the prediction at the point when training stops. Due to the early stopping logic, training stops after 50 iterates in which performance on the loss function on the test set has not improved.

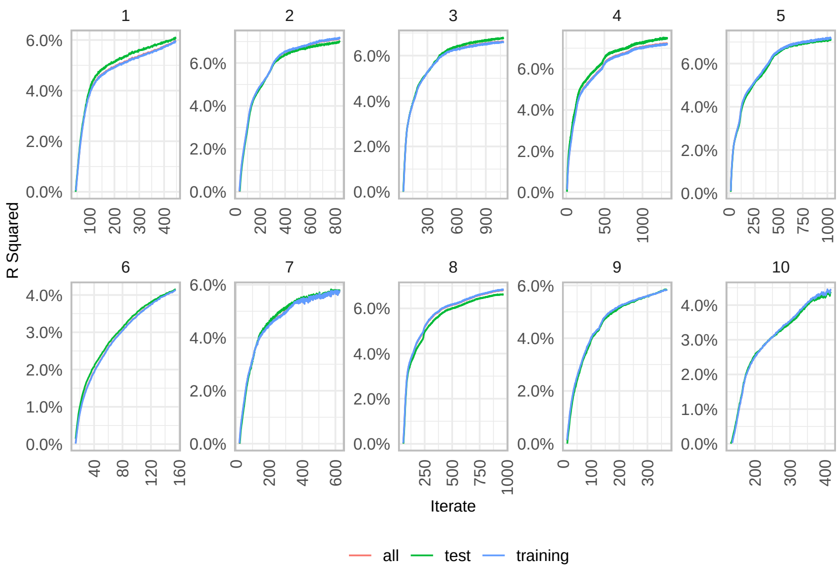
Table (J.19.2) Proportion of invalid values for each trained model.

Seed	Demand > 1	Demand < 0	Cost > Bid	Cost < 0
Combined	0.0002%	0.0002%	0.003%	0%
1	0%	0%	1.0064%	0%
2	0.0003%	0.0003%	0.7075%	0%
3	0%	0%	0.2606%	0%
4	0.0009%	0.0009%	0%	0%
5	0.0005%	0.0005%	0.025%	0%
6	0.0007%	0.0007%	1.7118%	0%
7	0.0003%	0.0003%	0.0016%	0%
8	0%	0%	0%	0%
9	0%	0%	0.2734%	0%
10	0.0002%	0.0002%	0.3094%	0%

Note: This table shows fraction of demands that are outside the unit interval and the fraction of costs that are above the final bid cost or below zero. Each figure in this table is a percentage of the non-missing values in the dataset. Figures are expressed to 4 decimal places.

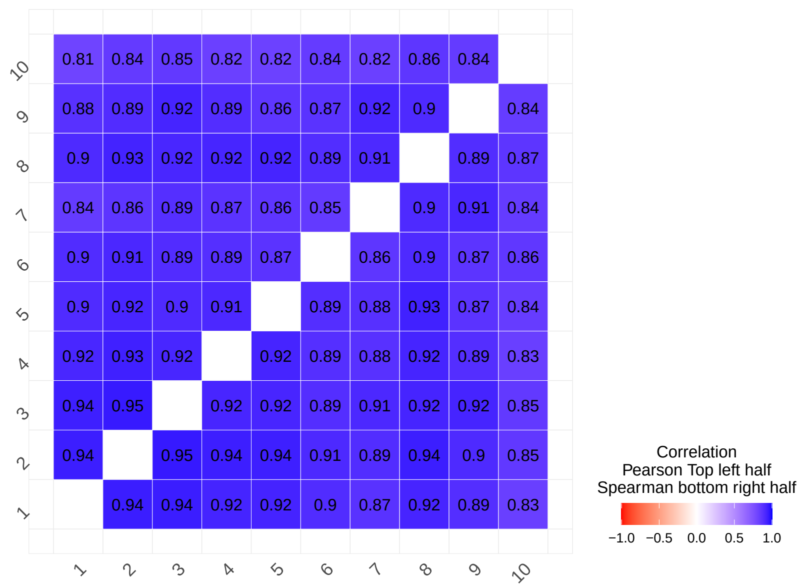
We first present tables and figures summarising the convergence path and fit of the model. The convergence plot of each seed with this specification can be seen in Figure J.19.1, while the correlation matrix between implied costs generated with each seed is displayed in Figure J.19.2. We summarise the

Figure (J.19.1) Convergence Plot for Specification 19



Note: This chart plots the convergence of each trained model in terms of the R^2. We only plot positive R^2 for each which is why some seeds appear to start later on. In general, R^2 values start training at negative values.

Figure (J.19.2) Correlation Matrix Between Costs Implied by Each Seed for Specification 19



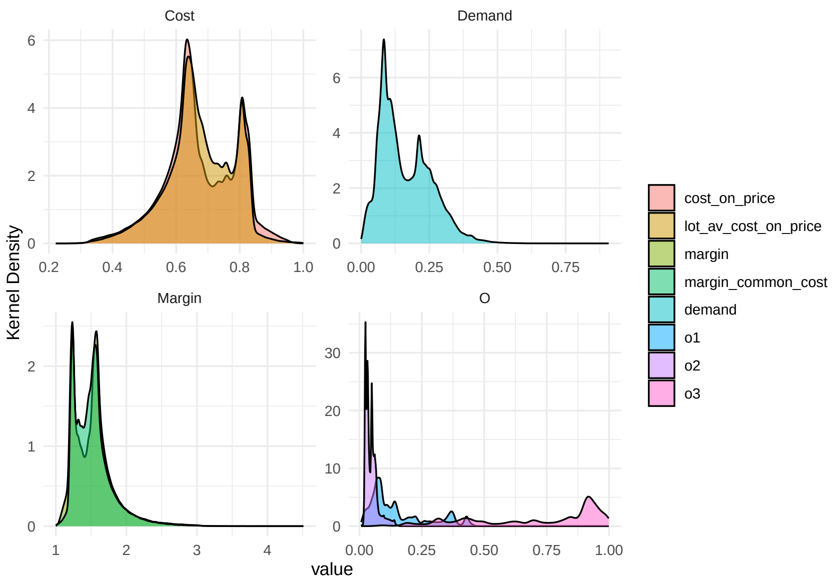
Note: This chart plots the correlation between $\frac{\text{Implied cost}}{\text{final price bid}}$ using the implied costs from each trained model. The top left half of the matrix are Pearson correlations while the bottom right half are Spearman correlations.

Table (J.19.3) Variable Importance Plot

Permuted Variable	Relative MSE	Relative Spearman Correlation
Price	15.0646	0.6661
Best Preceding Price	32.685	0.6994
Log Expected Price	1.0092	0.9402
Num. Bidders after	1.0636	0.5159
Mean prices bidder	1.0784	0.574
Num. Participants	1.0476	0.7218
Prop. Inactive Bidders	1.0543	0.777
Price Reduction in Round 3	1.0018	0.9891
Round 2 std of bids	1.1317	0.4785
Year tender published	1.0031	0.9819
Seller Size	1.0009	0.9953
Days Since Business Registered	1.0004	0.998

Note: To construct this table we took the 5 trained models (of 10 in total) with above median performance on the test set. We recorded it's performance on the full (test + training) sample as a benchmark. We then permuted each of the input variables (30 times per model) and recorded their performance. We recorded the relative MSE and Spearman correlation for each permutation and each model. We then took the medians for each model. Finally we took the median across models to get one relative performance metric per permuted variable. A relative MSE above 1 indicates a deterioration in performance by permuting that variable while a relative spearman below 1 indicates the same.

Figure (J.19.3) Kernel Densities of Values Estimated with Specification 19



Note: This chart plots the kernel densities of demand, implied margin, cost and the o_1, o_2, o_3 parameters. For the purposes of this figure, we remove observations where $\frac{\text{Implied cost}}{\text{final price bid}}$ is outside the $[0, 1]$ range or demand is outside this same range. The amount of observations that are dropped in this way are small and represented in Table J.19.2.

final convergences achieved by each seed in Table J.19.1. The fraction of demands and costs that fall outside the feasible range is in Table J.19.2. Note that models that do not converge well may have more estimated probabilities outside the unit interval. Our analysis only uses the 5 trained models that have the highest R^2 however these poorly performing cases are not used. The variable importance of each of the model's input variables is summarised in Table J.19.3. We then present kernel densities of demand, cost, margin, o_1, o_2 and o_3 in Figure J.19.3. In Table J.19.4 (corresponding to Table 4) we provide the regression coefficients of margin with different fixed effects for winners and losers.

Table (J.19.4) The Year-End Effect for Margins (with Varying Fixed Effects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Winners								
Last month	2.82***	2.96***	2.56***	2.34***	2.6***	2.41***	2.39***	2.4***
	(0.467)	(0.448)	(0.441)	(0.417)	(0.442)	(0.414)	(0.44)	(0.441)
R^2	0.00205	0.0271	0.319	0.159	0.321	0.177	0.436	0.438
N	112103	112099	112103	112103	112099	112099	112103	112099
Panel B: Losers								
Last month	4.08***	3.97***	3.45***	3.53***	3.48***	3.45***	3.16***	3.18***
	(0.611)	(0.578)	(0.522)	(0.52)	(0.522)	(0.511)	(0.483)	(0.483)
R^2	0.00192	0.0314	0.245	0.141	0.248	0.161	0.344	0.346
N	324096	324088	324096	324096	324088	324088	324096	324088
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	No	Yes	No	No	Yes	Yes	No	Yes
Seller FE	No	No	Yes	No	Yes	No	Yes	Yes
Buyer FE	No	No	No	Yes	No	Yes	Yes	Yes

Note: The sample is composed of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is a margin common cost as defined in equation 12. Last month is a dummy variable that takes the value one if the tender was finalised in December or November and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and by seller. *p<0.1; **p<0.05; ***p<0.01.

The regressions with the interaction terms are in Table J.19.5 (which corresponds to Table 5).

Finally in Table J.19.5 we replicate the regressions without controls on different time intervals of data. Specifically, for all bids, winners and losers respectively we perform separate regressions: one on 2017-2019 (pre-Covid period) and one on 2020-2021 (where the bidding process was impacted by covid). We present our findings in Table J.19.6.

Table (J.19.5) The Year-End Effect for Margins (with Varying Controls)

	All bids			*Winners*			*Losers*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Last month	2.79***	1.94***	2.35***	2.4***	1.58***	2.22***	3.18***	2.26***	2.58***
	(0.432)	(0.402)	(0.5)	(0.441)	(0.418)	(0.562)	(0.483)	(0.446)	(0.571)
Last month × 4 bidders			0.313			-0.29			0.515
			(0.679)			(0.796)			(0.727)
Last month × 5 bidders			0.147			-0.301			0.523
			(0.923)			(1.12)			(0.987)
Last month × 6 bidders			-1.87			-2.03			-1.65
			(1.16)			(1.34)			(1.24)
Last month × 7+ bidders			-2.74**			-4.19***			-2.7**
			(1.09)			(1.23)			(1.16)
4 bidders		-2.49***	-2.53***		-4.33***	-4.29***		-2.35***	-2.42***
		(0.247)	(0.261)		(0.268)	(0.288)		(0.272)	(0.289)
5 bidders		-6.59***	-6.6***		-9.73***	-9.68***		-6.59***	-6.66***
		(0.33)	(0.344)		(0.378)	(0.406)		(0.349)	(0.364)
6 bidders		-10.9***	-10.6***		-14.0***	-13.8***		-11.0***	-10.8***
		(0.421)	(0.443)		(0.485)	(0.521)		(0.449)	(0.47)
7+ bidders		-16.0***	-15.7***		-17.6***	-17.1***		-16.9***	-16.5***
		(0.422)	(0.432)		(0.488)	(0.497)		(0.45)	(0.459)
R^2	0.315	0.33	0.33	0.438	0.456	0.456	0.346	0.362	0.362
N	436187	436187	436187	112099	112099	112099	324088	324088	324088
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Columns 1, 4, and 7 correspond to specification 13. Columns 2, 5, and 8 also include dummies for each number of bidders in the auction. Columns 3, 6, and 9 further split the last month dummy by the number of bidders in the auction. Year FE stands for year fixed effects, buyer FE refers to government department fixed effects, seller FE indicates firm fixed effects, and sector FE represents sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

Table (J.19.6) The Year-End Effect for Margins (with Split by Year Interval)

	All 17-19	*All 20-21*	*Winners 17-19*	*Winners 20-21*	*Losers 17-19*	*Losers 20-21*
	(1)	(2)	(3)	(4)	(5)	(6)
Last month	2.81***	2.57***	2.62***	2.18***	3.12***	2.85***
	(0.54)	(0.574)	(0.638)	(0.599)	(0.61)	(0.652)
R^2	0.398	0.372	0.526	0.529	0.438	0.406
N	175872	260315	46371	65728	129501	194587
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation 12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

J.20 Specification 20

This is similar to the benchmark specification but the hidden layers are of width 2 rather than 4.

Table (J.20.1) Final convergence

		R^2			Pearson Corr			MAE		
Seed	Num. Iterates	All	Test	Training	All	Test	Training	All	Test	Training
1	741	0.0596	0.0618	0.0593	0.251	0.2565	0.2504	0.2498	0.2497	0.2498
2	523	0.045	0.0458	0.0449	0.2187	0.2196	0.2186	0.2644	0.2655	0.2642
3	500	0.054	0.0528	0.0541	0.2331	0.2307	0.2334	0.2564	0.2569	0.2563
4	738	0.0399	0.0386	0.04	0.2075	0.2041	0.2079	0.2539	0.2542	0.2539
5	1332	0.0648	0.0626	0.065	0.2555	0.2514	0.2559	0.2607	0.2609	0.2607
6	242	0.0301	0.0301	0.0301	0.1804	0.1815	0.1803	0.2715	0.2705	0.2716
7	1744	0.0512	0.053	0.0511	0.2283	0.232	0.2279	0.2519	0.2511	0.252
8	347	0.0255	0.0247	0.0256	0.1643	0.1615	0.1646	0.2755	0.2764	0.2754
9	824	-0.0074	-0.0077	-0.0073	0.0026	6e-04	0.0028	0.2695	0.2689	0.2696
10	1247	0.0632	0.0618	0.0633	0.2538	0.2503	0.2541	0.2511	0.2505	0.2512

Note: This table shows the R^2, Pearson correlation and Mean Absolute Error (MAE) of the prediction at the point when training stops. Due to the early stopping logic, training stops after 50 iterates in which performance on the loss function on the test set has not improved.

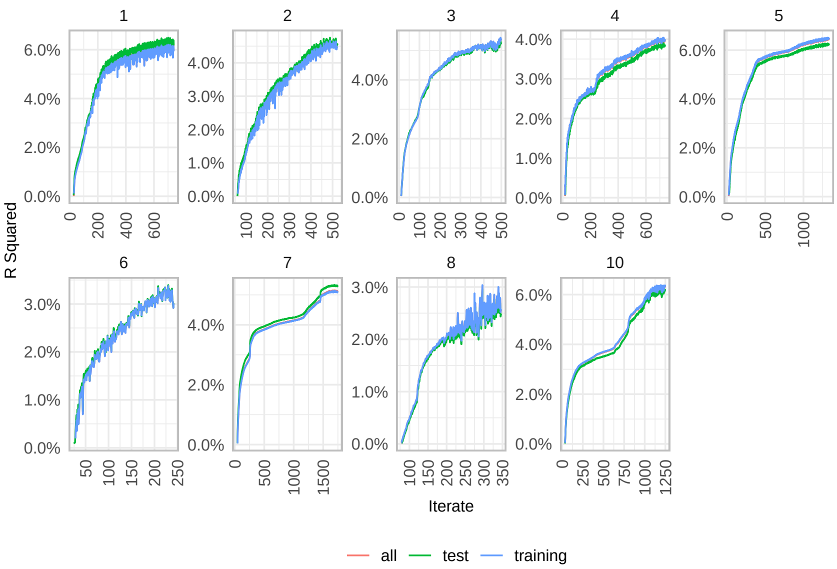
Table (J.20.2) Proportion of invalid values for each trained model.

Seed	Demand > 1	Demand < 0	Cost > Bid	Cost < 0
Combined	0%	0%	0%	0%
1	0%	0%	0%	0%
2	0.0012%	0.0012%	0%	0%
3	0.0001%	0.0001%	0%	0%
4	0%	0%	0%	0%
5	0%	0%	0%	0%
6	0%	0%	0.6472%	0%
7	0%	0%	0%	0%
8	0%	0%	0%	0%
9	0%	0%	0%	0%
10	0.0001%	0.0001%	0%	0%

Note: This table shows fraction of demands that are outside the unit interval and the fraction of costs that are above the final bid cost or below zero. Each figure in this table is a percentage of the non-missing values in the dataset. Figures are expressed to 4 decimal places.

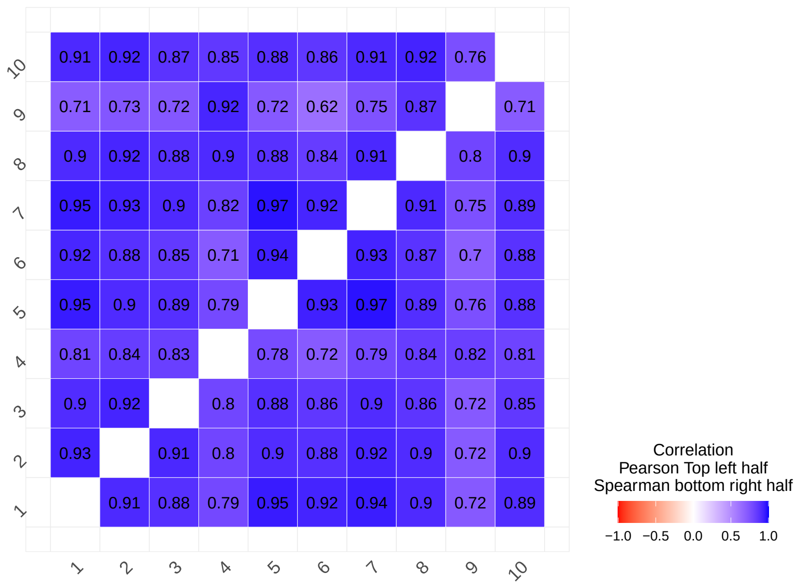
We first present tables and figures summarising the convergence path and fit of the model. The convergence plot of each seed with this specification can be seen in Figure J.20.1, while the correlation matrix between implied costs generated with each seed is displayed in Figure J.20.2. We summarise the final convergences achieved by each seed in Table J.20.1. The fraction of demands and costs that fall outside the feasible range is in Table J.20.2. Note that models that do not converge well may have more estimated probabilities outside the unit interval. Our analysis only uses the 5 trained models that have the highest R^2 however these poorly performing cases are not used. The variable importance of each of the model's input variables is summarised in Table J.20.3. We then present kernel densities of demand, cost, margin, o_1, o_2 and o_3 in Figure J.20.3. In Table J.20.4 (corresponding to Table 4) we provide the

Figure (J.20.1) Convergence Plot for Specification 20



Note: This chart plots the convergence of each trained model in terms of the R^2. We only plot positive R^2 for each which is why some seeds appear to start later on. In general, R^2 values start training at negative values.

Figure (J.20.2) Correlation Matrix Between Costs Implied by Each Seed for Specification 20



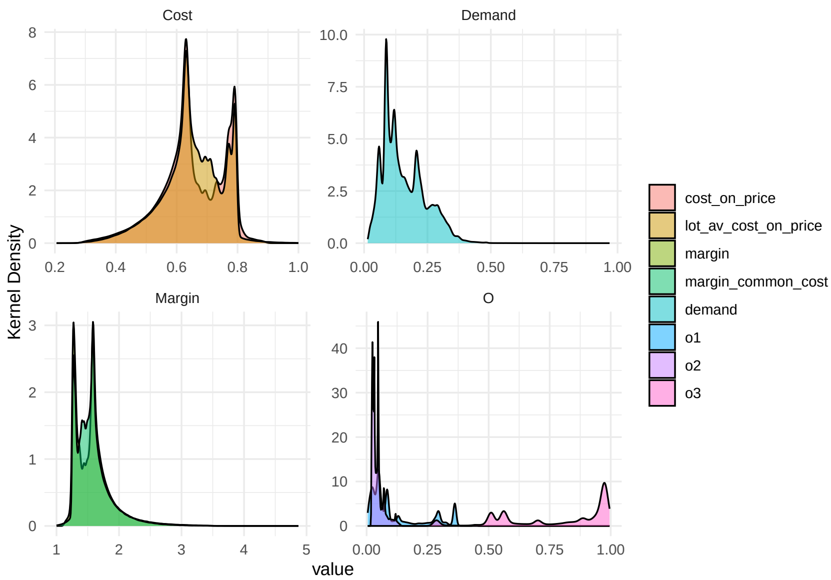
Note: This chart plots the correlation between $\frac{\text{Implied cost}}{\text{final price bid}}$ using the implied costs from each trained model. The top left half of the matrix are Pearson correlations while the bottom right half are Spearman correlations.

Table (J.20.3) Variable Importance Plot

Permuted Variable	Relative MSE	Relative Spearman Correlation
Price	7.5899	0.682
Best Preceding Price	22.0096	0.7238
Log Expected Price	1.0053	0.9555
Num. Bidders after	1.0537	0.5042
Mean prices bidder	1.0541	0.5904
Num. Participants	1.0268	0.7695
Prop. Inactive Bidders	1.0234	0.8472
Price Reduction in Round 3	1.0019	0.9874
Round 2 std of bids	1.122	0.4468
Year tender published	1.0025	0.9813

Note: To construct this table we took the 5 trained models (of 10 in total) with above median performance on the test set. We recorded it's performance on the full (test + training) sample as a benchmark. We then permuted each of the input variables (30 times per model) and recorded their performance. We recorded the relative MSE and Spearman correlation for each permutation and each model. We then took the medians for each model. Finally we took the median across models to get one relative performance metric per permuted variable. A relative MSE above 1 indicates a deterioration in performance by permuting that variable while a relative spearman below 1 indicates the same.

Figure (J.20.3) Kernel Densities of Values Estimated with Specification 20



Note: This chart plots the kernel densities of demand, implied margin, cost and the o_1, o_2, o_3 parameters. For the purposes of this figure, we remove observations where $\frac{\text{Implied cost}}{\text{final price bid}}$ is outside the $[0, 1]$ range or demand is outside this same range. The amount of observations that are dropped in this way are small and represented in Table J.20.2.

regression coefficients of margin with different fixed effects for winners and losers.

Table (J.20.4) The Year-End Effect for Margins (with Varying Fixed Effects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Winners								
Last month	2.6***	2.45***	2.01***	1.93***	2.11***	2.01***	1.91***	2.0***
	(0.423)	(0.414)	(0.406)	(0.377)	(0.405)	(0.377)	(0.401)	(0.401)
R^2	0.00124	0.0311	0.299	0.142	0.301	0.163	0.403	0.405
N	136756	136751	136756	136756	136751	136751	136756	136751
Panel B: Losers								
Last month	3.94***	3.56***	3.09***	3.22***	3.18***	3.12***	2.94***	3.03***
	(0.554)	(0.537)	(0.489)	(0.475)	(0.488)	(0.473)	(0.46)	(0.459)
R^2	0.00239	0.0371	0.231	0.126	0.234	0.15	0.319	0.322
N	382559	382550	382559	382559	382550	382550	382559	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	No	Yes	No	No	Yes	Yes	No	Yes
Seller FE	No	No	Yes	No	Yes	No	Yes	Yes
Buyer FE	No	No	No	Yes	No	Yes	Yes	Yes

Note: The sample is composed of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is a margin common cost as defined in equation 12. Last month is a dummy variable that takes the value one if the tender was finalised in December or November and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and by seller. *p<0.1; **p<0.05; ***p<0.01.

The regressions with the interaction terms are in Table J.20.5 (which corresponds to Table 5).

Finally in Table J.20.5 we replicate the regressions without controls on different time intervals of data. Specifically, for all bids, winners and losers respectively we perform separate regressions: one on 2017-2019 (pre-Covid period) and one on 2020-2021 (where the bidding process was impacted by covid). We present our findings in Table J.20.6.

Table (J.20.5) The Year-End Effect for Margins (with Varying Controls)

	All bids			Winners			Losers		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Last month	2.58***	1.73***	1.65***	2.0***	1.21***	1.55***	3.03***	2.09***	1.87***
	(0.404)	(0.376)	(0.471)	(0.401)	(0.383)	(0.522)	(0.459)	(0.425)	(0.539)
Last month × 4 bidders			0.613			0.17			0.823
			(0.645)			(0.77)			(0.693)
Last month × 5 bidders			1.15			-0.0744			1.48
			(0.858)			(1.06)			(0.907)
Last month × 6 bidders			-1.19			-2.44*			-0.849
			(1.04)			(1.25)			(1.11)
Last month × 7+ bidders			-1.36			-2.43**			-1.32
			(0.978)			(1.19)			(1.05)
4 bidders		-2.32***	-2.41***		-4.49***	-4.52***		-2.07***	-2.21***
		(0.239)	(0.257)		(0.268)	(0.29)		(0.265)	(0.284)
5 bidders		-6.82***	-6.99***		-10.3***	-10.3***		-6.75***	-6.96***
		(0.306)	(0.317)		(0.358)	(0.383)		(0.325)	(0.339)
6 bidders		-11.2***	-11.0***		-14.9***	-14.5***		-11.2***	-11.1***
		(0.394)	(0.419)		(0.476)	(0.507)		(0.421)	(0.446)
7+ bidders		-16.7***	-16.5***		-18.4***	-18.1***		-17.5***	-17.4***
		(0.414)	(0.428)		(0.506)	(0.516)		(0.443)	(0.457)
R^2	0.29	0.306	0.306	0.405	0.423	0.423	0.322	0.338	0.338
N	519301	519301	519301	136751	136751	136751	382550	382550	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Columns 1, 4, and 7 correspond to specification 13. Columns 2, 5, and 8 also include dummies for each number of bidders in the auction. Columns 3, 6, and 9 further split the last month dummy by the number of bidders in the auction. Year FE stands for year fixed effects, buyer FE refers to government department fixed effects, seller FE indicates firm fixed effects, and sector FE represents sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

Table (J.20.6) The Year-End Effect for Margins (with Split by Year Interval)

	All 17-19	*All 20-21*	*Winners 17-19*	*Winners 20-21*	*Losers 17-19*	*Losers 20-21*
	(1)	(2)	(3)	(4)	(5)	(6)
Last month	2.46***	2.36***	2.22***	1.59***	2.75***	2.76***
	(0.516)	(0.527)	(0.623)	(0.546)	(0.594)	(0.601)
R^2	0.374	0.34	0.5	0.489	0.415	0.373
N	208764	310537	56309	80442	152455	230095
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation 12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

J.21 Specification 21

This is similar to the benchmark specification but the hidden layers are of width 3 rather than 4.

Table (J.21.1) Final convergence

		R^2			Pearson Corr			MAE		
Seed	Num. Iterates	All	Test	Training	All	Test	Training	All	Test	Training
1	844	0.058	0.0611	0.0577	0.2455	0.2524	0.2447	0.2484	0.2482	0.2484
2	1255	0.0662	0.0657	0.0662	0.261	0.2608	0.2611	0.2495	0.2507	0.2493
3	396	0.0486	0.0469	0.0488	0.2322	0.2289	0.2325	0.2476	0.2482	0.2475
4	558	0.0581	0.0573	0.0582	0.2425	0.2407	0.2427	0.2549	0.2551	0.2549
5	903	0.0546	0.0538	0.0547	0.2439	0.2414	0.2441	0.2487	0.2487	0.2487
6	1249	0.0659	0.0645	0.0661	0.259	0.2555	0.2594	0.2521	0.2515	0.2522
7	763	0.0662	0.0671	0.0661	0.258	0.26	0.2578	0.2546	0.254	0.2547
8	793	0.0643	0.0637	0.0644	0.2549	0.2537	0.255	0.2528	0.2537	0.2527
9	620	0.06	0.0608	0.0599	0.2472	0.2487	0.2471	0.2521	0.2513	0.2522
10	1004	0.0656	0.0635	0.0658	0.2564	0.2524	0.2568	0.2588	0.2581	0.2588

Note: This table shows the R^2, Pearson correlation and Mean Absolute Error (MAE) of the prediction at the point when training stops. Due to the early stopping logic, training stops after 50 iterates in which performance on the loss function on the test set has not improved. Note that poorly converging models are not be used as for our analysis we take only the 5 models with the highest R^2.

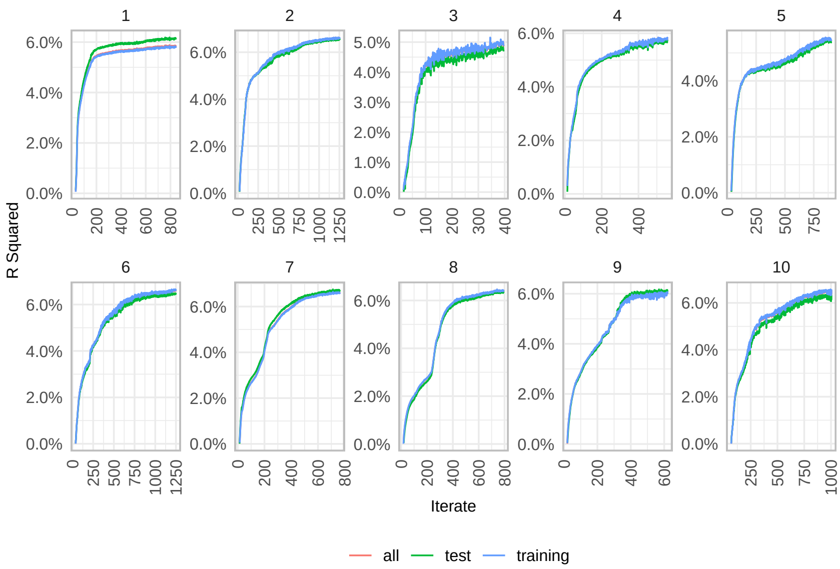
Table (J.21.2) Proportion of invalid values for each trained model.

Seed	Demand > 1	Demand < 0	Cost > Bid	Cost < 0
Combined	0.0009%	0.0009%	0%	0%
1	0%	0%	0%	0%
2	0.0007%	0.0007%	0%	0%
3	0%	0%	0%	0%
4	0%	0%	0.2162%	0%
5	0.0001%	0.0001%	0%	0%
6	0.0002%	0.0002%	0%	0%
7	0.0006%	0.0006%	0%	0%
8	0.001%	0.001%	0.0002%	0%
9	0.001%	0.001%	0%	0%
10	0.0044%	0.0044%	0%	0%

Note: This table shows fraction of demands that are outside the unit interval and the fraction of costs that are above the final bid cost or below zero. Each figure in this table is a percentage of the non-missing values in the dataset. Figures are expressed to 4 decimal places. Note that poorly converging models might have many invalid demands but they will not be used as for our analysis we take only the 5 models with the highest R^2.

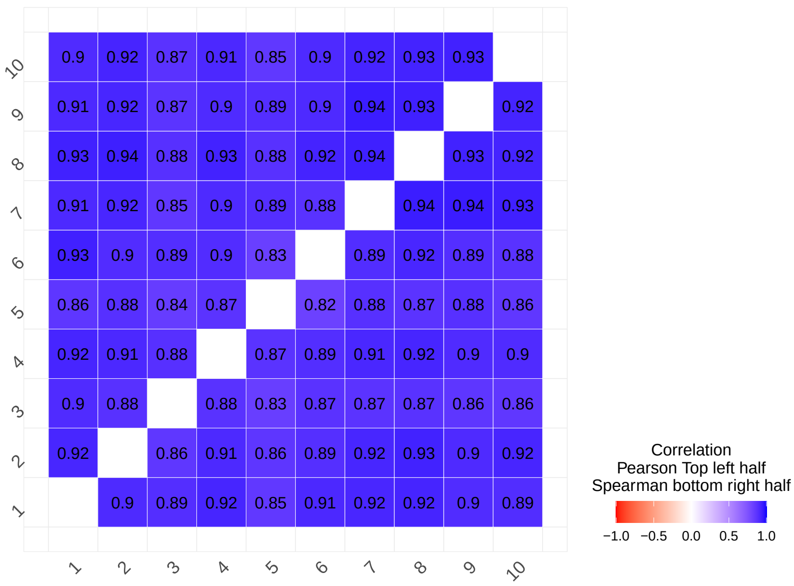
We first present tables and figures summarising the convergence path and fit of the model. The convergence plot of each seed with this specification can be seen in Figure J.21.1, while the correlation matrix between implied costs generated with each seed is displayed in Figure J.21.2. We summarise the final convergences achieved by each seed in Table J.21.1. The fraction of demands and costs that fall outside the feasible range is in Table J.21.2. Note that models that do not converge well may have more estimated probabilities outside the unit interval. Our analysis only uses the 5 trained models that have

Figure (J.21.1) Convergence Plot for Specification 21



Note: This chart plots the convergence of each trained model in terms of the R^2. We only plot positive R^2 for each which is why some seeds appear to start later on. In general, R^2 values start training at negative values.

Figure (J.21.2) Correlation Matrix Between Costs Implied by Each Seed for Specification 21



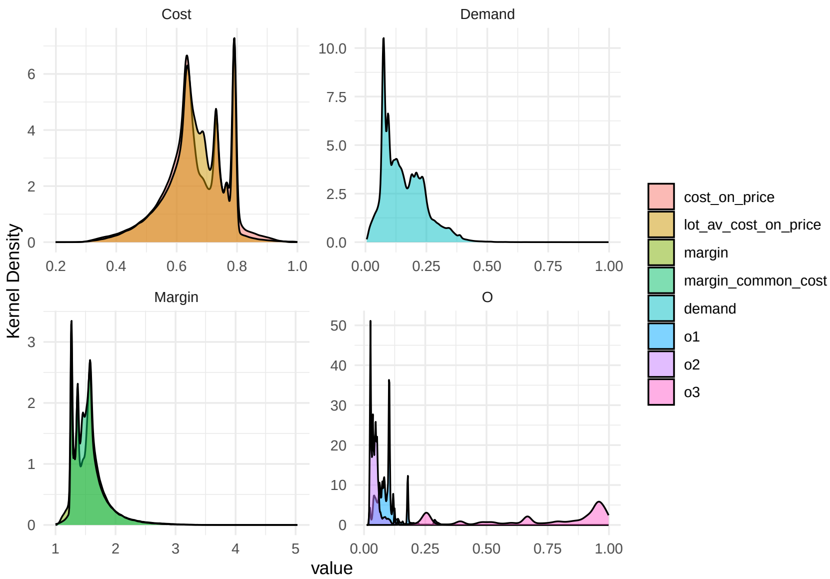
Note: This chart plots the correlation between $\frac{\text{Implied cost}}{\text{final price bid}}$ using the implied costs from each trained model. The top left half of the matrix are Pearson correlations while the bottom right half are Spearman correlations.

Table (J.21.3) Variable Importance Plot

Permuted Variable	Relative MSE	Relative Spearman Correlation
Price	11.2382	0.6652
Best Preceding Price	8.7007	0.7143
Log Expected Price	1.0053	0.9653
Num. Bidders after	1.0503	0.5395
Mean prices bidder	1.0659	0.5771
Num. Participants	1.0317	0.8325
Prop. Inactive Bidders	1.0402	0.7899
Price Reduction in Round 3	1.0015	0.9901
Round 2 std of bids	1.1334	0.4163
Year tender published	1.0026	0.9818

Note: To construct this table we took the 5 trained models (of 10 in total) with above median performance on the test set. We recorded it's performance on the full (test + training) sample as a benchmark. We then permuted each of the input variables (30 times per model) and recorded their performance. We recorded the relative MSE and Spearman correlation for each permutation and each model. We then took the medians for each model. Finally we took the median across models to get one relative performance metric per permuted variable. A relative MSE above 1 indicates a deterioration in performance by permuting that variable while a relative spearman below 1 indicates the same.

Figure (J.21.3) Kernel Densities of Values Estimated with Specification 21



Note: This chart plots the kernel densities of demand, implied margin, cost and the o_1, o_2, o_3 parameters. For the purposes of this figure, we remove observations where $\frac{\text{Implied cost}}{\text{final price bid}}$ is outside the $[0, 1]$ range or demand is outside this same range. The amount of observations that are dropped in this way are small and represented in Table J.21.2.

the highest R^2 however these poorly performing cases are not used. The variable importance of each of the model's input variables is summarised in Table J.21.3. We then present kernel densities of demand, cost, margin, o_1, o_2 and o_3 in Figure J.21.3. In Table J.21.4 (corresponding to Table 4) we provide the regression coefficients of margin with different fixed effects for winners and losers.

Table (J.21.4) The Year-End Effect for Margins (with Varying Fixed Effects)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Winners								
Last month	2.53***	2.37***	1.97***	1.91***	2.06***	1.96***	1.88***	1.96***
	(0.403)	(0.391)	(0.384)	(0.359)	(0.384)	(0.357)	(0.381)	(0.381)
R^2	0.00147	0.0294	0.295	0.143	0.297	0.162	0.4	0.402
N	136756	136751	136756	136756	136751	136751	136756	136751
Panel B: Losers								
Last month	3.91***	3.5***	3.09***	3.24***	3.17***	3.1***	2.95***	3.03***
	(0.535)	(0.513)	(0.468)	(0.459)	(0.468)	(0.454)	(0.442)	(0.442)
R^2	0.0018	0.0329	0.225	0.124	0.228	0.145	0.313	0.316
N	382559	382550	382559	382559	382550	382550	382559	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	No	Yes	No	No	Yes	Yes	No	Yes
Seller FE	No	No	Yes	No	Yes	No	Yes	Yes
Buyer FE	No	No	No	Yes	No	Yes	Yes	Yes

Note: The sample is composed of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is a margin common cost as defined in equation 12. Last month is a dummy variable that takes the value one if the tender was finalised in December or November and zero otherwise. Year FE stands for year fixed effects, buyer FE are government departments fixed effects, firm FE are firm fixed effects, and industry FE are sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and by seller. *p<0.1; **p<0.05; ***p<0.01.

The regressions with the interaction terms are in Table J.21.5 (which corresponds to Table 5).

Finally in Table J.21.5 we replicate the regressions without controls on different time intervals of data. Specifically, for all bids, winners and losers respectively we perform separate regressions: one on 2017-2019 (pre-Covid period) and one on 2020-2021 (where the bidding process was impacted by covid). We present our findings in Table J.21.6.

Table (J.21.5) The Year-End Effect for Margins (with Varying Controls)

	All bids (1)	(2)	(3)	*Winners* (4)	(5)	(6)	*Losers* (7)	(8)	(9)
Last month	2.58***	1.69***	1.53***	1.96***	1.17***	1.41***	3.03***	2.05***	1.74***
	(0.387)	(0.358)	(0.441)	(0.381)	(0.362)	(0.487)	(0.442)	(0.406)	(0.506)
Last month × 4 bidders			0.612			0.155			0.838
			(0.619)			(0.735)			(0.663)
Last month × 5 bidders			1.24			0.178			1.57*
			(0.83)			(1.02)			(0.879)
Last month × 6 bidders			-0.873			-2.08*			-0.492
			(1.01)			(1.19)			(1.07)
Last month × 7+ bidders			-1.11			-1.91*			-1.11
			(0.941)			(1.13)			(1.01)
4 bidders		-1.24***	-1.33***		-3.4***	-3.42***		-1.01***	-1.14***
		(0.229)	(0.246)		(0.256)	(0.277)		(0.253)	(0.272)
5 bidders		-5.62***	-5.8***		-9.03***	-9.05***		-5.55***	-5.79***
		(0.302)	(0.312)		(0.349)	(0.372)		(0.321)	(0.333)
6 bidders		-10.5***	-10.4***		-14.1***	-13.8***		-10.6***	-10.5***
		(0.383)	(0.408)		(0.457)	(0.489)		(0.413)	(0.436)
7+ bidders		-17.5***	-17.4***		-18.7***	-18.5***		-18.4***	-18.3***
		(0.412)	(0.426)		(0.486)	(0.499)		(0.443)	(0.458)
R^2	0.285	0.303	0.303	0.402	0.422	0.422	0.316	0.336	0.336
N	519301	519301	519301	136751	136751	136751	382550	382550	382550
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Columns 1, 4, and 7 correspond to specification 13. Columns 2, 5, and 8 also include dummies for each number of bidders in the auction. Columns 3, 6, and 9 further split the last month dummy by the number of bidders in the auction. Year FE stands for year fixed effects, buyer FE refers to government department fixed effects, seller FE indicates firm fixed effects, and sector FE represents sectoral fixed effects. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

Table (J.21.6) The Year-End Effect for Margins (with Split by Year Interval)

	All 17-19 (1)	*All 20-21* (2)	*Winners 17-19* (3)	*Winners 20-21* (4)	*Losers 17-19* (5)	*Losers 20-21* (6)
Last month	2.4***	2.43***	2.13***	1.6***	2.7***	2.85***
	(0.494)	(0.505)	(0.597)	(0.518)	(0.572)	(0.578)
R^2	0.371	0.334	0.498	0.486	0.412	0.368
N	208764	310537	56309	80442	152455	230095
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Buyer FE	Yes	Yes	Yes	Yes	Yes	Yes
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: The sample consists of winners and losers of competitive lots for which we estimated costs. The dependent variable in all columns is the margin common cost, as defined in equation 12. "Last month" is a dummy variable that takes the value of one if the tender was finalized in December or November and zero otherwise. Standard errors are in parentheses and are clustered by buyer and seller. *p<0.1; **p<0.05; ***p<0.01.

References

Bandiera, Oriana, Andrea Prat, and Tommaso Valletti. 2009. "Active and Passive Waste in Government Spending: Evidence from a Policy Experiment." *American Economic Review*, 99(4): 1278–1308.

Baumann, Stuart. 2019. "Putting It Off for Later: Procrastination and End-of-Fiscal-Year Spending Spikes." *The Scandinavian Journal of Economics*, 121(2): 706–735.

Firpo, Sergio, Nicole M. Fortin, and Thomas Lemieux. 2009. "Unconditional Quantile Regressions." *Econometrica*, 77(3): 953–973.

Guerre, Emmanuel, Isabelle Perrigne, and Quang Vuong. 2000. "Optimal Nonparametric Estimation of First-Price Auctions." *Econometrica*, 68(3): 525–574.

Hyndman, Noel, Rowan Jones, and Maurice Pendlebury. 2007. "An Exploratory Study of Annuality in the UK Public Sector: Plus ça change, plus c'est la même chose." *Financial Accountability and Management*, 23(2): 0267–4424.

Kawai, Kei, and Jun Nakabayashi. 2022. "Detecting Large-Scale Collusion in Procurement Auctions." *Journal of Political Economy*, 130(5): 1364–1411.

Liebman, Jeffrey, and Neale Mahoney. 2017. "Do Expiring Budgets Lead to Wasteful Year-End Spending? Evidence from Federal Procurement." *American Economic Review*, 107(11): 3510–3549.

Nakabayashi, Jun. 2013. "Small Business Set-asides in Procurement Auctions: An Empirical Analysis." *Journal of Public Economics*, 100: 28–44.

Open Government Partnership. 2016. "Open Government Awards." *https://www.opengovpartnership.org/2016-open-government-awards/*.

Open Government Partnership. 2021. "Open Government Awards." *https://www.opengovpartnership.org/open-government-awards/*.

Perrigne, Isabelle, and Quang Vuong. 2019. "Econometrics of Auctions and Nonlinear Pricing." *Annual Review of Economics*, 11: 27–54.

Procurement Leaders. 2016. "Public Sector: Ministry of Economic Development and Trade, Ukraine." *https://events.procurementleaders.com/events/awards/world-procurement-awards?tab=2017-winners*.

Ramachandran, Prajit, Barret Zoph, and Quoc Le. 2017. "Searching for Activation Functions." *arXiv preprint arXiv:1710.05941*.

State Procurement Agency of Georgia. 2015. "Georgian Successful Reforms in Public Procurement: Path Towards the Risk-Based and IT Based Approach." *World Bank Group*.

Transparency International. 2012. "Corruption Perceptions Index 2020." *https://www.transparency.org/en/press/20121205-transparency-international-georgia-*

51st-in-2012-corruption-percept.

Transparency International Ukraine. 2017. "Co-Creation of ProZorro: An Account of The Process and Actors." *https://www.transparency.org/en/publications/co-creation-of-prozorro-an-account-of-the-process-and-actors*.

Verkhovna Rada. 2015. "The Law on Public Procurement." *https://zakon.rada.gov.ua/laws/show/922-19/ed20151225*.

World Bank Group. 2018. "Improving Efficiency in Public Procurement in Georgia. Findings from Data Analysis of Public Procurement Transactions in Georgia 2013-2016."