This is a post-print version prepared for Researchgate of
Alfnes, F., Rickertsen, K. and Shogren, J. F. (2017), Test-Retesting in Experimental Valuation of Perishable Food Products: Unstable Individual Bids and Reliable Market Demand. Journal of Agricultural Economics doi:10.1111/1477-9552.12248

Test-Retesting in Experimental Valuation of Perishable Food Products: Unstable Individual Bids and Reliable Market Demand

Frode Alfnes, Kyrre Rickertsen and Jason F. Shogren[1]

Abstract

One important property of a preference measure is its reliability. In this article, we explore the reliability of experimental auctions in measuring the market demand for five types of fish. We use the test-retest method to compare demand curves from two Becker-DeGroot-Marschak experiments with the same 116 participants conducted seven months apart in time. The individual bids are not stable for these perishable products, but the distributions of the individual bids are stable. We find that the unsystematic individual variations cancel out in the aggregation of bids in a typical sample size for experimental valuation studies. Our results suggest that experimental auctions provide reliable market demand estimates even though the individual bids may change substantially over time.

Keywords: *Test-retest; reliability; aggregation; BDM mechanism.*
JEL classifications: *C91, D12, Q13*.

---

[1] Frode Alfnes and Kyrre Rickertsen are both with the School of Economics and Business, Norwegian University of Life Sciences, Norway. Email: frode.alfnes@nmbu.no for correspondence. Jason Shogren is in the Department of Economics and Finance, University of Wyoming. The paper was written while Jason Shogren was adjunct professor at the School of Economics and Business, Norwegian University of Life Sciences. Pierre Combris at Institut National de la Recherche Agronomique (INRA), Ivry-sur-Seine, France, and Sylvie Issanchou and Géraldine Enderli at Centre des Sciences du Goût et de l'Alimentation, CNRS, INRA, University of Bourgogne Franche-Comté, Dijon, France helped us with collecting data and provided valuable comments. Their help is highly appreciated. The Research Council of Norway, grants 178300/I10 and 199564/I10 provided financial support for this research.

1. Introduction

Experimental valuation methods such as experimental auctions, real choice experiments and price list experiments are often used to measure preferences for goods and services (e.g. Lusk and Shogren, 2007; Alfnes and Rickertsen, 2011). These methods have in common that consumers have an opportunity to buy products in the experiments; and for food products this usually means consuming it within hours or days after the purchase. This makes experimental valuation sensitive to short-term variations in preferences due to factors such as purchase intentions on the day of the experiment (Corrigan and Rousu, 2008), how hungry the participants are at the time of the experiment (de-Magistris and Gracia, 2016), possibly what they have eaten in the days before the experiment or what they already have in their fridge. These time-varying individual factors are likely most important in valuation of perishable products such as fresh fish, in which consumption need to take place within a short period after the experiment.

Several studies report instabilities in the elicited individual preferences. Mattei (2000), for example, reports preference reversals and unstable individual preferences within one laboratory experiment session. Furthermore, unstable preferences as reflected by increasing bids within one experimental session have frequently been reported in multiple-trial Vickrey auctions (e.g. Shogren *et al.*, 1994; Alfnes and Rickertsen, 2003). Increasing bids within one session may be explained by phenomena like preference learning (e.g. List and Shogren, 1999) or bid affiliation (e.g. Corrigan and Rousu, 2006).

Non-informative signals in the form of arbitrary anchors may also influence behaviour in experiments (e.g. Ariely *et al.*, 2003). For example, Nunes and Boatwrigth (2004) show how prices of an unrelated product seen outside the valuation area affects the valuation of a product. Others have studied how factors related to the participants and that typically change over time, affects the valuation. De Magistris and Gracia (2016), for instance, found that degree of hunger mattered for the valuation of food products even when the products would be eaten a week later.

Individual preference changes are an important topic in behavioural economics (e.g. Tversky and Thaler, 1990). However, for most food valuation purposes the reliability of aggregated market demand is what matters most. The large majority of papers using experimental valuation methods focus on the effect of product attributes, information, or sociodemographic or psychological segmentation variables on aggregate demand (means, medians, demand curves) (Lusk and Shogren, 2007; Alfnes and Rickertsen, 2011).

By reliability, we mean the consistency of the measure over time. In psychometrics, reliability is frequently assessed by using the test-retest method (e.g. Groth-Marnat, 2009). This method has also been used in stated preference studies to examine the stability of hypothetical statements of value for recreation demand (e.g. Loomis, 1989). In the case of experimental valuation methods, the test-retest method implies that people who participate in an experiment are brought back into the lab at a future date and asked to repeat the valuation exercise. As far as we know, only Shogren *et al.* (2000) have used the test-retest approach in experimental valuation methods. They examined the reliability of values for irradiated pork over a two-week window using experimental auctions and observed that mean bids remained relatively stable for the well-known goods, but declined for the novel good.

The stability of preferences has been tested in previous studies on market data, and in a food demand context, the focus has been on specific food categories like beef, poultry, pork and fish. This earlier work assumed the existence of a representative consumer and tested the stability of preferences using aggregate time series data over a long period of time. Stable preferences at the market level have been both rejected (e.g. Moschini and Mielke, 1989; Sakong and Hayes, 1993; Cortez and Senauer, 1996; Gallet and List, 1998; Jin and Koo, 2003) and not rejected (e.g. Chalfant and Alston, 1988; Burton and Young, 1991). Reported changes in market preferences have typically been attributed to changes in external factors such as health information, advertising, or food scares. In these studies, it is usually not discussed to what extent stable market preferences are a result of aggregation across consumers.

Stable market demand as a result of aggregation is reported in Härdle and Kirman (1995) who studied professional buyers in the Marseilles fish market. They found downward sloping demand curves at the aggregate level but not, in general, at the individual level, and summarised their result as 'sophisticated and complicated individual behavior may lead to simple aggregate properties'.

To investigate the reliability of aggregated demand measures from experimental valuation methods, we look at stability in preferences[2] at both the individual and the market level.

[2] The textbook case of stability in preferences is that an individual has a unique utility function over all available products. Changes in willingness-to-pay for a product will then be a result of changes in income or changes in prices of the other products. An alternative interpretation of stability in preferences is stability

Experimental auctions such as the Becker-DeGroot-Marschak (BDM) mechanism (Becker *et al.*, 1964) gives individual bids representing the participants willingness-to-pay for the products on offer, while other experimental valuation methods such as the real choice experiment or price list experiments use the distribution of choices to estimate a willingness-to-pay distribution. Since only the auctions give individual willingness-to-pay directly, they are better suited to study the relationship between variability in individual preferences and reliability in market demand curves than other experimental valuation methods. Stable individual willingness-to-pay is a sufficient, but not a necessary condition for a stable market demand curve. As shown by Becker (1962), well-behaved downward sloping demand curves can be derived from a wide class of behaviour, including random choice within the budget set. The presence of unsystematic individual specific effects would also make market behaviour more regular than individual behaviour since these effects will cancel out by aggregation.

We investigate to what extent experimental valuation methods generate reliable market demand for perishable products. We use a panel of 116 French consumers who participated in two experimental auctions conducted seven months apart for five types of fish.

2. Test-Retest Experimental Auctions: Sample and Design

of some 'underlying preferences' as discussed by Becker (1976). In his own words (p. 5): 'The preferences that are assumed to be stable do not refer to market goods and services, like oranges, automobiles, or medical care, but to underlying objects of choice that are produced by each household using market goods and services, their own time and other inputs. These underlying preferences are defined over fundamental aspects of life, such as health, prestige, sensual pleasure, benevolence or envy, that do not always bear a stable relation to market goods and services'. Another alternative interpretation of stability in preferences, which is highly relevant for food markets, is stability of 'food values' such as safety, nutrition and taste as discussed by Lusk and Briggeman (2009). Both of these alternative interpretations of stability in preferences may explain why people buy different foods on different days, but they do not explain how much variation can be expected in the valuation of a market product. Neither, do they say anything about the reliability of experimental valuation methods in eliciting individual bids or aggregate demand for specific products.

The data used in this test-retest come from a consumer study on fish presented in Rickertsen *et al.* (2017). A series of experiments were conducted with a panel of French consumers in the sensory laboratory of I'institut National de la Recherche Agronomique (INRA) in Dijon in 2008. A sample of 180 fish eating consumers was recruited by phone from INRA's consumer panel for an experiment conducted in May. To test for stability and reliability, all participants were contacted again seven months later, and 116 of the 180 came back for a new experiment in December. We use these 116 participants as our sample in this paper. The sample consisted of 116 consumers from Dijon, 63 women and 53 men, of ages ranging from 23 to 70 years old, with an average age of 48 years and standard deviation of 12.4. A majority of 71 were working full or part time, 2 were students and the remaining 44 were pensioners, homemakers or unemployed. All the participants said they are part of the food decisions in their household, eat fish at least once a month and purchase fish at least every second month. For our purpose of testing of stability and reliability, the important feature of the sample is that the same 116 subjects participated both times.[3]

Between 12 and 16 people participated in each session. At each of the two sessions, the participants were paid €25 and each participant evaluated five types of fish: salmon, wild cod, farmed cod, monkfish and pangasius. Each session had two parts and a survey. The first part was a sensory evaluation. After a professional chef cooked the fish, each participant was served 50 grams of each fish in rotation to avoid any ordering effects. Before tasting, the participants were told what fish they were served. After tasting, the participants gave their hedonic scores on a computer.

The second part of the experiment elicited preferences for the fish. We used a BDM mechanism, where the participants bid on a product and their bids are compared to a randomly drawn price. If their bids are lower than the price, they do not buy. If the bids are higher than the price, they pay the price and get the product. Each package of fish weighed about 300 grams, which may be insufficient for a meal in a household consisting of more than two persons. To avoid

[3] The 116 May participants that showed up again in December had a slightly higher mean WTP for the five fishes than the 64 that did not, 8.31 vs 7.68 Euros. Using a two-sample Wilcoxen rank-sum test of differences in median bids between the two groups, we only find a significant difference for wild cod (salmon: z= –0.23, p = 0.82; wild cod: –2.21, p = 0.03; farmed cod: z = –1.38, p = 0.17; monkfish z = –0.15, *p* = 0.88; pangasius z = –0.23, *p* = 0.82).

package size causing zero bids, we let each participant choose if he or she wanted to purchase 1, 2, 3, 4 or 5 packages of fish before the bidding. The separation of bid and price means that each bidder's weakly dominant strategy is to bid the amount that leaves him or her indifferent between obtaining the product or not. The participants gave a bid on each of the five types of fish on a computer. Next, one product and one price was randomly drawn as binding. All the participants that had bid higher than the price were allowed to buy the fish drawn as binding. Due to this procedure, no one bought more than one type of fish and those that had bid less than the price for the binding fish did not buy fish. Although the BDM mechanism is weakly demand revealing in theory, these mechanisms require initial training (e.g. Lusk and Shogren, 2007, p. 63). Our training had two parts. First we explained the BDM and second, the participants practiced by bidding on orange juice or champagne. The training auctions were hypothetical and no products were sold.

As part of the survey conducted after the BDM, we asked the respondents about their 'best guess at the average market price for one kilogram of fresh salmon, cod, monkfish and pangasius fillets this week'. They answered on a 12-point scale starting at €3 and increasing with €3 all the way up to €36. The median price guess for salmon and cod was unchanged at €12 and €15 between the two sessions, and monkfish decreased from €21 to €18 and pangasius decreased from €9 to €7.5. Using a Wilcoxon matched-pairs signed-ranks test, the decline in the perceived market price for monkfish is significant ($z=1.96$, $p=0.05$), while the others have no significant changes.

The recruitment of participants, the size of the groups, the lab, the software, the training, the BDM valuation method and the random drawing of product and price were all established procedures used at INRA's sensory lab in Dijon both before and after our experiment.

3. Results

Table 1 shows the descriptive statistics for the test-retest bids. The median bids remained reliable from May to December for all fish types except farmed cod. For farmed cod, the median bid increased from €8.00 to €9.25. The mean bid for farmed cod increased from €7.82 to €8.71 while the mean bids fell slightly for the other fish types.

Insert Table 1 here

3.1. Reliability results

Two main results on bid stability and aggregated reliability emerge from our test-retest experiments.

Result 1. BDM experiments result in unstable individual willingness-to- pay across two sessions conducted months apart.

Support. In Row 1 of Table 2, we present the number of participants with constant bids (mostly 0 in both experiments). Rows 2 through 4 show the number of participants who changed bids by less than half a standard deviation; by one-half to one standard deviation; and by more than one standard deviation. Table 1 shows that the standard deviations range from €3.74 to €6.91 for the different fishes. More than 40% of the participants changed their bids by more than half a standard deviation for salmon, wild and farmed cod and monkfish. Pangasius, with 73% zeros, had 24% that changed their bids with more than half a standard deviation.

Many participants also changed their ranking of the five fish types. Only 9% of the participants had a constant ranking for all five fish types. A constant ranking means that if they bid higher for one type of fish than another in May, they also did that in December. Furthermore, as shown by the last row of Table 2, only 46% of the participants ranked salmon identically according to the bids in May and December. The corresponding percentages for wild cod, farmed cod, monkfish and pangasius are 41, 37, 53 and 59, respectively.

Table 3 shows the Pearson correlation coefficients between the bids in May and December and the associated 95% confidence intervals. Zero correlation suggests independent valuations in the two experiments, i.e. random bidding, while a correlation factor of one implies constant bidding. All the correlation coefficients are significantly different from zero and one, suggesting that the individual bids are positively correlated between the two experiments.

Insert Table 2 here

Insert Table 3 here

Result 2. The BDM mechanism is a reliable method for eliciting aggregate market demand curves across two sessions conducted months apart.

Support. Define each individual's demand curve by:

$$(1) \qquad x_{ijt} = \begin{cases} 0 \ if \ p_{jt} > Bid_{ijt} \\ 0 \ if \ p_{jt} \leq Bid_{ijt} \end{cases}$$

where x_{ijt} is the quantity demanded of product *j* by individual *i* at time *t*, p_{jt} is the price of the good, and Bid_{ijt} is the bid.[4] Let participant 1 be the highest bidder, participant 2 the second-highest bidder, and so on; *n* participants are willing to pay at least the same as participant *n*'s bid. The market demand curves, X_{jt} is defined by:

$$(2) \qquad X_{jt} = \sum_{i=1}^{I} x_{ijt}(p_{jt}).$$

We constructed the market demand curves for May using equation (2), i.e. by ranking each individual participant's bid from the highest to the lowest. The market demand curves for salmon, wild cod and farmed cod are shown by the solid lines in Panels A, C and E of Figure 1. In these panels, the December bids using the same ordering of the participants as in May are plotted as the dots.

As discussed above, the individual bids changed in erratic ways. The changes are likely to be the result of day-to-day variations due to factors such as what people have eaten in the days before the experiment. Examples of the erratic bidding can be found in Figure 1. Panel A shows 11 participants who bid a positive amount for salmon in May, but bid zero in December; 10 participants who bid zero in December, but bid a positive amount in May. Panels C and E show similar erratic changes in bids for wild and farmed cod. For example, the second highest bidder

[4] We let each participant choose if he or she wanted to purchase 1, 2, 3, 4 or 5 packages of fish before the bidding. Since this procedure was implemented for practical reasons, we count the individual demand in equation (1) as 1 rather than the requested number of packages.

for wild cod in May, who bid €20, was only the 56th highest bidder in December with €10, and the third highest bidder in December, who bid €19.90, was only the 29th highest bidder in December with €14. Of the 11 participants who bid a positive amount for wild cod in May, but not in December, 9 bid a positive amount for salmon in December. This result shows that even though these participants did not want to buy wild cod, they still wanted to buy fish.

To investigate the stability of market demand, we constructed the market demand curves in December as in May, i.e. by ranking the participants bids from the highest to the lowest bid. Panels B, D and F present the results for salmon, wild cod and farmed cod. As shown in the figures, none of the market demand curves shift substantially,[5] but for farmed cod we can see a minor shift from May to December.

Several statistical tests were implemented to test for the stability of the market demand curves. We use a Wilcoxon signed rank test (W) to test for identical median bids, a Brown-Forsythe test (B-F) to test for identical variances of bids, and a Kolmogorov-Smirnov test (K-S) to test for identical probability distribution functions for bids in the two experiments (Hollander and Wolfe, 1999). Table 3 presents the test values and the associated *p*-values of the tests. As indicated by the shift for farmed cod seen in Figure 1, the median bid for farmed cod increased significantly. This increase could be explained by improved quality as reflected by a significant increase in the median hedonic score. The median bids for the other fish types did not change. Finally, we do not reject identical variance or identical distribution functions for any of the bids in May and December.

Insert Figure 1 here

4. Concluding Remarks

The reliability of a measure is an important property. An ideal measure would have a low level of random errors (high reliability), control for alternative causal explanations (high internal validity), and give results that can be generalised to other populations and conditions (high external validity).

In this article, we explore the stability and reliability of experimental auctions in eliciting preferences at the individual and market level for a perishable product. We use a test-retest experimental design eliciting values seven months apart from French consumers for five types of

[5] Figures for monkfish and pangasius show similar patterns.

fish. Since fresh fish is a perishable product, fish bought in the experiment ought to be consumed within few days after the experiment. Hence, day-to-day variations in preferences are likely to affect the WTP for the product in the experiment. Our results show that individual bids were unstable over the experimental sessions. However, we find that these unsystematic individual variations cancel out in the aggregation of bids in a typical sample size for experimental valuation studies. That a reliable measure for market preferences can emerge due to the aggregation across consumers, supports the broader rationality notion that reliability in economic valuation is better defined as a social construct, rather than an individual one (Arrow, 1987; Smith, 2003).

Our laboratory results support the general empirical notion that random day-to-day variations in individual preferences have minor effects on the stability and reliability of the predicted market demand. This is important for food producers and retailers who are primarily interested in the market demand and not the preferences of each individual consumer.

For researchers interested in segmentation, the unsystematic individual variations means that the sociodemographic sub-samples must be large enough for the aggregation to cancel out the unsystematic individual variations. A remaining challenge for users of experimental valuation auctions is to demonstrate to what degree the experimental methods, with their rigour with respect to internal validity, also are able to predict behaviour in non-experimental markets (also see Shogren *et al.*, 1999).

References

Alfnes, F. and Rickertsen, K. ‘European consumers’ willingness to pay for U.S. beef in experimental auction markets’, *American Journal of Agricultural Economics,* Vol. 85, (2003) pp. 396–405.

Alfnes, F. and Rickertsen, K. ‘Non-market valuation: experimental methods’, in Lusk, J., Roosen, J. and Shogren, J.F. (eds.) *The Oxford Handbook of The Economics of Food Consumption and Policy* (Oxford: Oxford University Press. 2011).

Ariely, D., Lowenstein, G. and Prelec, D. ‘Coherent arbitrariness: Stable demand curves without stable preferences’, *Quarterly Journal of Economics,* Vol. 118, (2003) pp. 73–104.

Arrow, K. ‘Rationality of self and others in an economic system’, in Hogarth, R. and Reder, M. (eds.) *Rational Choice: The Contrast between Economics and Psychology* (Chicago, IL: University of Chicago Press, 1987).

Becker, G.S. ‘Irrational behavior and economic theory’, *The Journal of Political Economy,* Vol. 70, (1962) pp. 1–13.

Becker, G.S. *The Economic Approach to Human Behavior* (Chicago, IL: The University of Chicago Press, 1976).

Becker, G.M., DeGroot, M.H. and Marschak. J. ‘Measuring utility by a single response sequential method’, *Behavioral Science,* Vol. 9, (1964) pp. 226–232.

Burton, M.P. and Young, T. ‘Non-parametric tests for changes in consumer preferences for meat in Great Britain’, *Journal of Agricultural Economics,* Vol. 42(2), (1991), pp. 138–145.

Chalfant, J.A. and Alston, J.M. ‘Accounting for changes in tastes’, *Journal of Political Economy,* Vol. 96, (1988) pp. 391–410.

Corrigan, J.R. and Rousu, M.C. ‘Posted prices and bid affiliation: Evidence from experimental auctions’, *American Journal of Agricultural Economics,* Vol. 88, (2006) pp. 1078–1090.

Corrigan, J.R. and Rousu, M.C. ‘Testing whether field auction experiments are demand revealing in practice’, *Journal of Agricultural and Resource Economics,* Vol. 33, (2008) pp. 290–301.

Cortez, R. and Senauer. B. ‘Taste changes in the demand for food by demographic groups in the United States: a nonparametric empirical analysis’, *American Journal of Agricultural Economics,* Vol. 78, (1996) pp. 280–289.

De-Magistris, T. and Gracia, A., ‘Assessing projection bias in consumers’ food preferences’, *PloS one*, Vol. 11(2), (2016), p.e0146308.

Gallet, C.A. and List. J.A. ‘Elasticities of beer demand revisited’, *Economics Letters,* Vol. 61, (1998) pp. 67–71.

Groth-Marnat, G. *Handbook of Psychological Assessment*, 5th edition (Hoboken, NJ: John Wiley & Sons, 2009).

Härdle, W. and Kirman, A. ‘Nonclassical demand: a model-free examination of price-quantity relations in the Marseille fish market’, *Journal of Econometrics,* Vol. 67, (1995) pp. 227–257.

Hollander, M. and Wolfe D. *Nonparametric Statistical Methods*, 2nd edition. (New York: Wiley, 1999).

Jin, H.J. and Koo, W.W. ‘The effects on the BSE outbreak in Japan on consumer preferences’, *European Review of Agricultural Economics,* Vol. 30, (2003) pp. 173–192.

List, J.A. and Shogren, J.F. ‘Price information and bidding behavior in repeated second-price auctions’, *American Journal of Agricultural Economics,* Vol. 81 (1999) pp. 942–949.

Loomis, J. ‘Test-retest reliability of the contingent valuation method: a comparison of general population and visitor responses’, *American Journal of Agricultural Economics,* Vol. 71, (1989) pp. 76–84.

Lusk, J.L. and Briggeman, B.C. ‘Food values’, *American Journal of Agricultural Economics,* Vol. 91, (2009) pp. 184–196.

Lusk, J.L. and Shogren, J.F. *Experimental Auctions: Methods and Applications in Economic and Marketing Research* (Cambridge: Cambridge University Press, 2007).

Mattei, A. ‘Full-scale real test of consumer behavior using experimental data’, *Journal of Economic Behavior & Organization,* Vol. 43, (2000) pp. 487–497.

Moschini, G. and Mielke, K.D. ‘Modeling the pattern of structural change in U.S. meat demand’, *American Journal of Agricultural Economics,* Vol. 71, (1989) pp. 253–261.

Nunes, J.C. and Boatwright, P. 'Incidental prices and their effect on willingness to pay', *Journal of Marketing Research*, Vol. 41, (2004) pp. 457–466.

Rickertsen, K., Alfnes, F., Combris, P., Enderli, G., Issanchou, S. and Shogren, J.F. 'French consumers' attitudes and preferences toward wild and farmed fish', *Marine Resource Economics*, Vol. 32, (2017) pp. 59–81.

Sakong, Y. and Hayes. D.J. 'Testing the stability of preferences: a nonparametric approach', *American Journal of Agricultural Economics,* Vol. 78, (1993) pp. 280–289.

Shogren, J.F., Hayes, D.J., Kliebenstein, J.B., and Fox, J.A. 'Bid sensitivity and the structure of the Vickrey auction', *American Journal of Agricultural Economics*, Vol. 76, (1994) pp. 1089–1095.

Shogren, J.F., Fox, J.A., Hayes, D.J., and Roosen, J. 'Observed choices for food safety in retail, survey, and auction markets', *American Journal of Agricultural Economics*, Vol. 81, (1999) pp. 1192–1199.

Shogren, J.F., List, J.A., and Hayes, D.J. 'Preference learning in consecutive experimental auctions', *American Journal of Agricultural Economics,* Vol. 82, (2000) pp. 1016–1021.

Smith, V.L. 'Constructivist and ecological rationality in economics', *American Economic Review,* Vol. 93, (2003) pp. 465–508.

Tversky, A. and Thaler, R.H. 'Preference reversals', *Journal of Economic Perspectives,* Vol. 4, (1990) pp. 193–205.

Table 1

Descriptive statistics for the bids in euro. Sample size=116

		Mean	Median	St Dev	Min	Max	#Zero bids
Salmon							
	May	10.16	10.00	5.70	0.00	25.00	15
	December	10.01	10.00	5.10	0.00	18.00	13
Wild cod							
	May	9.87	10.00	5.59	0.00	23.00	16
	December	9.81	10.00	5.66	0.00	20.00	18
Farmed cod							
	May	7.82	8.00	5.50	0.00	21.00	25
	December	8.71	9.25	5.79	0.00	22.00	23
Monkfish							
	May	11.40	12.00	6.60	0.00	25.00	20
	December	10.75	12.00	6.91	0.00	24.00	24
Pangasius							
	May	2.28	0.00	4.06	0.00	18.00	82
	December	1.96	0.00	3.74	0.00	17.00	87

Note: The price guesses (Min, Median, Max) in May and December were (6,12,18) and (6,12,21) for salmon, (6,15,24) and (6,15,27) for cod, (12, 21, 36) and (12, 18, 33) for monkfish, and (3,9, 18) and (3, 7.5, 15) for pangasius, respectively. We did not specify if the fish was farmed or wild in the price question.

Table 2

Stability of individual bids (in %)

	Salmon	Wild Cod	Farmed Cod	Monkfish	Pangasius	Total
Constant bid	18	15	15	19	68	
0 < Change < 0.5·SD[a]	38	27	32	35	9	
0.5·SD ≤ Change ≤ 1.0·SD[a]	27	25	22	21	8	
Change > 1.0·SD[a]	17	34	31	26	16	
Constant ranking	46	41	37	53	59	9

Note: [a] SD is standard deviation.

Table 3

Test results for reliability

	P[a]	95% CI[b]	W[c]	p-value[d]	B-F[e]	p-value[d]	K-S[f]	p-value[d]
Salmon	0.63	0.50-0.73	0.16	0.87	1.20	0.28	0.08	0.88
Wild cod	0.46	0.31-0.59	-0.07	0.95	0.01	0.94	0.03	1.00
Farmed cod	0.51	0.36-0.63	-2.12	0.03	0.55	0.46	0.12	0.37
Monkfish	0.53	0.39-0.65	0.62	0.53	1.03	0.31	0.09	0.57
Pangasius	0.65	0.54-0.75	0.33	0.74	1.18	0.37	0.40	0.53

Notes:

[a] Pearson correlation coefficient as estimated by STATA corrci command.

[b] 95 % confidence interval for Pearson correlation coefficient as estimated by STATA corrci command using the Fisher transformation.

[c] Wilcoxon signed-rank test as estimated by STATA signrank command.

[d] The p-value of the test statistic reported in the previous column.

[e] Brown-Forsythe test of equal variance as estimated by STATA robvar command.

[f] Kolmogorov–Smirnov test as estimated by STATA ksmirnov command.

Figure 1. Test-retest bids for salmon and wild cod in May and December