ARTICLE IN PRESS

Journal of Econometrics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Assessing consumer demand with noisy neural measurements

Ryan Webb [a,*,1], Nitin Mehta [a], Ifat Levy [b]

[a] Rotman School of Management, University of Toronto, Canada
[b] School of Medicine, Yale University, United States of America

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Neuroeconomics
Measurement error
Choice prediction
Valuation
Proxy variables

ABSTRACT

Recent studies have used the random utility framework to examine whether neural data can assess and predict demand for consumer products, both within and across individuals. However the effectiveness of this methodology has been limited by the large degree of measurement error in neural data. The resulting "error-in-variables" problem severely biases the estimates of the relationship between neural measurements and choice behaviour, thus limiting the role such data can play in assessing marginal contributions to utility. In this article, we propose a method for controlling for this large degree of measurement error in value regions of the brain. We propose that additional neural variables from areas of the brain that are unrelated to valuation can serve as "proxies" for the measurement error in value regions, substantially alleviating the bias in model estimates. We demonstrate the feasibility of our proposed method on an existing dataset of fMRI measurements and consumer choices. We find a substantial reduction in the bias of model estimates compared to existing baseline methods (the estimated coefficients roughly double), leading to improved inference and out-of-sample demand prediction. After controlling for measurement error, we also find a considerable reduction in the variation of model estimates across consumers.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Given a sample of consumer choice data and other observables, establishing and measuring the marginal contribution to utility of these observables is now standard practice in economic and marketing analysis (McFadden, 2001). These methods allow a researcher to assess the effect of a product or policy change on consumer demand, predict future demand, and potentially infer the resulting impact on consumer welfare (McFadden, 2013).

One challenge in discrete choice analysis lies in measuring the demand for goods that are not currently provided in the market. One possible approach is to estimate counterfactual demand using parametric assumptions based on the marginal utilities of attributes of existing goods. However, without a full understanding of how attributes contribute to utility, such assumptions are quite strong for predicting the elasticity of demand for novel products (e.g. wearable computers or flying taxis). For this reason, responses to hypothetical surveys are typically used instead; the evaluation or design of novel products typically relies on *conjoint analysis*, in which stated-preferences are used to assess and aggregate hypothetical attributes. Similarly, the assessment of environmental goods relies on stated preference methods for *contingent valuation*.

* Corresponding author.
E-mail address: ryan.webb@rotman.utoronto.ca (R. Webb).

[1] The authors gratefully acknowledge discussions with V. Aguirregabiria, A. Goldfarb, I. Mourifié, M. Osborne, D. Soberman, and seminar participants at the University of California — Berkeley. R.W. acknowledges funding from a SSHRC Insight Development grant 430-2019-00246. N.M. acknowledges funding from a SSHRC Insight grant 435-2019-0062. I.L. acknowledges funding from NSF grant BCS-1829439.

https://doi.org/10.1016/j.jeconom.2020.07.028
0304-4076/© 2020 Elsevier B.V. All rights reserved.

Please cite this article as: R. Webb, N. Mehta and I. Levy, Assessing consumer demand with noisy neural measurements. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.07.028.

Recently, a growing literature has explored using additional neurobiological measurements to augment survey methods and improve demand prediction. It is now well-documented that signals in the human brain correlate with representative utilities measured while humans evaluate and make decisions over a wide range of choice objects (e.g. consumer goods, money lotteries, charitable donations, durable goods, and social and political outcomes; Levy and Glimcher, 2012; Bartra et al., 2013; Clithero and Rangel, 2013).[2] These correlations suggest a method for eliciting preferences when standard revealed preference data are noisy, problematic, or absent: measurements of neural activity can be used to estimate valuations directly, thus serve as explanatory variables to predict choice behaviour (Smith et al., 2014; Webb et al., 2019). Recent studies have demonstrated improved prediction results within individual,[3] across individuals,[4] and even for market outcomes like the effectiveness of advertising and crowd-funding campaigns (Falk et al., 2012; Venkatraman et al., 2015; Boksem and Smidts, 2015; Genevsky et al., 2017).

The analysis in nearly all of these studies can be framed in terms of the Random Utility Model (RUM) familiar to discrete choice analysts. Let X_j denote a typical vector of observables (e.g. brand dummies or product attributes) that enter the utility function for alternative j. Estimating the marginal utility (or hedonic weight), $\boldsymbol{\beta}$, of these observables via,

$$U_j = X_j\boldsymbol{\beta} + \epsilon_j, \tag{1}$$

with choice probabilities over alternatives given by,

$$\begin{aligned} P_j(X_j) &= \Pr[U_j > U_k, \quad \forall k \neq j] \\ &= \Pr[(X_j - X_k)\boldsymbol{\beta} > \epsilon_k - \epsilon_j, \quad \forall k \neq j], \end{aligned} \tag{2}$$

is the key step in predicting demand (McFadden, 2001).

This is also the case for recent neuroeconomic methods, with the distinction that observables also include neural measurements of the value that subjects place on brands or attributes (McFadden, 2013). The simplest form of these models specify utility directly in terms of the neural activations for an alternative (denoted N_j):

$$U_j = N_j\boldsymbol{\beta} + \epsilon_j. \tag{3}$$

The parameter vector $\boldsymbol{\beta}$ now captures the change in latent utility in response to measured neural activity for different choice alternatives (Knutson et al., 2007; Smith et al., 2014; Genevsky and Knutson, 2015; Genevsky et al., 2017; Webb et al., 2019). By far the most popular method for making neural measurements is functional Magnetic Resonance Imaging (fMRI), with previous studies relating fMRI measurements to a change in a product's price or packaging (Knutson et al., 2007; Lusk et al., 2016), the branding of a product (McClure et al., 2004), a change in the quantity of a product (Levy and Glimcher, 2011), the valuation of environmental goods (Khaw et al., 2015), or even the introduction of a novel product that a consumer has never experienced (Barron et al., 2013).

Despite the wide range of applications, one issue in particular has hampered the usefulness of neuroeconomic methods in assessing and predicting consumer choice. There is typically a large degree of error inherent in neural measurement techniques; an initial estimate of its standard deviation is over four times larger than the error in utility ϵ_j (Webb et al., 2019). This measurement error can arise from many sources, which we will discuss below, but its implications are stark. Consider the estimation of the simple model (3), but instead of directly observing neural activity, we only observe a noisy measurement of it. Denote this noisy signal $B_j = N_j + \mu_j$, where μ_j represents the measurement error. Then the utility specification becomes

$$\begin{aligned} U_j &= B_j\boldsymbol{\beta} + (\epsilon_j - \mu_j\boldsymbol{\beta}) \\ &= B_j\boldsymbol{\beta} + \tilde{\epsilon}_j. \end{aligned} \tag{4}$$

The error term $\tilde{\epsilon}_j$ from this model is now negatively correlated with the regressor of interest because of the relation between B_j and μ_j. This "error-in-variables" problem will not only decrease the precision of estimated marginal effects ($\boldsymbol{\beta}$), but also bias them towards zero (Yatchew and Griliches, 1985). This bias severely limits the ability of neuroeconomic techniques to assess the impact of variation in observables: it will worsen predictive performance and limit inference about whether which brain regions respond to a particular manipulation of observables. Indeed, measurement error will considerably raise the probability that a false null hypothesis might fail to be rejected, increasing Type II errors on hypotheses about the role of a brain region in decision-making.

In this article, we propose a new method for alleviating the error-in-variables problem in the context of neural data. This method will allow us to more accurately measure and predict the relationship between neural data and behavioural outcomes within the RUM. The intuition for our proposed method is as follows. Note that measurement error enters utility in (4) as an omitted variable. Therefore additional observables might exist which can act as a *proxy* for measurement error

[2] The utilities implied by revealed-preference methods (Hsu et al., 2009; Levy et al., 2011), willingness-to-pay methods (Plassmann et al., 2007), and stated-preference methods (Hare et al., 2010) have all been used as explanatory variables to isolate value signals in neural data.

[3] Knutson et al. (2007), Lebreton et al. (2009), Tusche et al. (2010), Levy et al. (2011), Smith et al. (2014), Gross et al. (2014), Telpaz et al. (2015) and Webb et al. (2019).

[4] For example, Smith et al. (2014), Gross et al. (2014) and Telpaz et al. (2015).

Please cite this article as: R. Webb, N. Mehta and I. Levy, Assessing consumer demand with noisy neural measurements. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.07.028.

(i.e. an observable that is correlated with measurement error, but would otherwise have no impact on utility). If such proxy variables can be found, then adding these variables to the utility specification (4) will reduce the bias in estimates of $\boldsymbol{\beta}$. We therefore propose using contemporaneous signals from regions of the brain unrelated to consumer preference as proxies for the measurement error in the regressors for utility. The rationale for these proxy variables lies in the properties of measurement error in fMRI data. As we will discuss in Section 2, the measurement error in value regions of the brain is still correlated with activity in brain regions unrelated to choice behaviour. Therefore a typical fMRI dataset — which contains contemporaneous measurements from other brain regions with no known relationship to valuation — might also contain information on both the direction and magnitude of the measurement error in value regions on any given measurement trial. If so, such contemporaneous measurements could be used as a proxy for unobserved measurement error to achieve more accurate and precise estimates of the relation between the neural data and choice behaviour.

We demonstrate our method using a previously-reported dataset containing fMRI measurements and binary choices over consumer products (Levy et al., 2011). We find a substantial reduction in the bias of model estimates compared to existing baseline methods (the estimated coefficients roughly double), leading to improved inference and out-of-sample demand prediction. After controlling for measurement error, we also find a considerable reduction in the variation of estimates across consumers. We also demonstrate that two standard methods for dealing with the error-in-variables problem, (i) averaging over additional independent measurements of B_j, and (ii) using additional independent measurements as instrumental variables, are not feasible in this context. This highlights how our proposed correction differs from the standard IV approach to the measurement error problem; we propose controlling for μ_j directly rather than hunting for an instrument that is correlated with B_j but not μ_j. Finally, our proposed method does not require any changes to existing experimental protocols (such as repeated choices from the same set, or calibration measurements), and in principle, is applicable to other measurement techniques such as Electroencephalography (EEG; Telpaz et al., 2015).

Our proposed error-in-variables correction is relevant to the growing econometric literature which relates neural observables to choice behaviour within the RUM framework, but which has not yet corrected for the error-in-variables problem. In a landmark study, Smith et al. (2014) assess the predictive ability of neural data for choice behaviour both within and across consumers. In an effort to guard against over-fitting they use data from all brain regions, but apply a LASSO estimator to shrink *all* the estimates of parameters in $\boldsymbol{\beta}$. While this approach nullifies the influence of many extraneous measurements, the resulting estimates are biased by definition. In contrast, the method proposed here uses regions specified *a priori*, with additional covariates included to control for measurement error without artificially shrinking the estimate of $\boldsymbol{\beta}$.

So far we have discussed our proposed methodology in the context of Eq. (3), in which the observables in utility include neural measurements. Our proposed methodology will also apply when the marginal utility of an observable attribute $x_{j,l} \in X_j$ is modelled explicitly (Harris and Keane, 1999).[5] For example, Lusk et al. (2016) specified the marginal utility of an attribute, β_l, as a function of measured neural activity such that $\beta_l \equiv \beta_{0,l} + \beta_{1,l} B_{j,l}$. These marginal effects are then jointly estimated with standard observables via:

$$\begin{aligned} U_j &= X_j(\boldsymbol{\beta}_0 + B_j\boldsymbol{\beta}_1) + \epsilon_j \\ &= X_j\boldsymbol{\beta}_0 + X_jB_j\boldsymbol{\beta}_1 + \epsilon_j. \end{aligned} \tag{5}$$

Note that this approach effectively introduces the neural observables as interaction terms with the other regressors in X_j (e.g. price levels or brand dummies), either directly (Lusk et al., 2016) or in two stages (Venkatraman et al., 2015).[6] These studies find considerable variation in neural signals across consumers which yields improved choice prediction results beyond the standard model. However, given a large degree of measurement error, these estimates of the interaction term coefficient are still biased towards zero. The method proposed here can alleviate this bias.[7]

Webb et al. (2019) have previously assessed the impact of measurement error in neural observables using data from the Levy et al. (2011) experiment. In this dataset, neural activity is measured for each alternative, then a subject makes repeated choices over all binary pairs of alternatives. Therefore the same error-ridden neural measurements can be used to compare repeated choices of the same pair of alternatives, yielding an estimate of the standard deviation of the measurement error and a partial bias reduction in the estimate of $\boldsymbol{\beta}$.[8] By contrast, the method proposed here directly controls for the error-in-variables problem, therefore provides an improved estimate of both $\boldsymbol{\beta}$ and the measurement error. Perhaps most importantly, our proposed method does not require any particular properties of the experimental

[5] Harris and Keane (1999) use survey responses (Likert scales) on the importance of attributes in healthcare plans to specify marginal effects. They found that including these scales significantly improved model fit. The relative performance of neural measures, perhaps in conjunction with simple Likert scales, remains to be explored.

[6] To see the relation to (3), consider a study which only varies choice alternatives. Then X_j is an alternative-specific dummy variable and $U_j = \alpha_j + N_j\boldsymbol{\beta}_1 + \epsilon_j$. We consider the relation between this model and (3) further in Section 5.3.

[7] For example, Lusk et al. (2016) did not observe a significant main effect of a change in mPFC activity in response to a change in price, in contrast with earlier literature (Knutson et al., 2007; Plassmann et al., 2008; Karmarkar et al., 2015). Since their study did not address the issue of measurement error, it is possible that the biased estimate of this relation led to a Type II error. The method proposed here can yield a less-biased, more precise estimate of this relation.

[8] Formally, measurement error can be modelled as a random effect which is constant over repetitions of the same pair but varies over pairs. Webb et al. (2019) apply this model assuming that measurement error is independent across pairs.

Please cite this article as: R. Webb, N. Mehta and I. Levy, Assessing consumer demand with noisy neural measurements. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.07.028.

design like repeated choice trials. It is thus widely applicable, both to the large number of existing datasets in the literature and to future studies.

Section 2 discuss background on the fMRI measurements used in the neuroeconomics literature and the dataset used in our study. Section 3 presents some initial reduced-form results to motivate our choice of proxy variables. Section 4 presents our modelling framework and the technical properties of the measurement error correction. Section 5 applies our proposed method to the Levy et al. (2011) dataset and compares with existing methods. Section 6 concludes.

2. Background

2.1. fMRI measurements of valuation

In the neuroeconomics literature, fMRI imaging of the Blood Oxygenation Level Dependent (BOLD) signal is the predominant method for measuring neural activity. This method takes separate measurements in each of about 100,000 27 mm^3 cubes (called *voxels*) tiling the human brain. This relatively high spatial resolution allows researchers to identify which regions of the brain (clusters of voxels) exhibit activity that positively correlates with the value of alternatives. Meta-analyses of over 200 studies demonstrate that the BOLD signal from primarily two clusters, located in the medial Pre-Frontal Cortex (mPFC) and the ventral Striatum (vSTR), correlate with the valuation of choice objects (Levy and Glimcher, 2012; Bartra et al., 2013; Clithero and Rangel, 2013).

While meta-analyses clarify the properties of the BOLD signal in mPFC and vSTR collected over many measurements, subjects, and studies, any single BOLD measurement from these regions is highly variable. One source of this variability arises simply because of the imperfect correlation between a neuron's activity and the electromagnetic properties measured by BOLD. More invasive methods for measuring neural activity (with electrodes implanted in a brain) reveal that the BOLD signal is far more variable than the underlying neural activity (Logothetis, 2002).

However the spatial and highly-distributed functional structure of the brain complicates this issue further. Not only will the activity within each brain region be measured with error, but contemporaneous errors in the BOLD signal are likely to be correlated across regions for both physiological and mechanical reasons. And perhaps more fundamentally, at the functional-level, different regions may be correlated with value regions due to some other functional role taking place alongside valuation. For example, visual cortex might respond to visual properties of the same stimuli evaluated by value regions. And value regions, such as the mPFC, can be engaged in many functions, such as social cognition and learning to inhibit maladaptive responses (Delgado et al., 2016). These relationships between regions can induce contemporaneous correlations that are not related to valuation, and can appear stable across trials, tasks, and individuals (Gratton et al., 2018).

This correlation across regions also presents on opportunity. In an evaluation of consumer behaviour, the aim is to understand how the measured BOLD signal changes as the subject evaluates different products or attributes. Clearly, measurement error makes this task more difficult. But, in principle, measurements from brain regions that are known to be uncorrelated with value (say, via large meta-studies) might provide information about the direction and magnitude of measurement error in mPFC and vSTR. Therefore any correlation between these regions and the mPFC or vSTR on a given measurement trial can be used as a proxy to control for the measurement error within mPFC or vSTR on that same trial. Even the measurement from a single voxel could be used in this role, provided it was uncorrelated with valuation and correlated enough with the measurement error in the contemporaneous signals from mPFC or vSTR. In this study, we consider measurements from voxels in two clusters which previous research suggests are unrelated to valuation.

Occipotal cortex. The first cluster is in the occipital cortex (OCC), in the vicinity of primary visual cortex. This is the same control region examined in Levy et al. (2011), who verified that activity in this region was not correlated with the rank ordering of choices. This region was limited to voxels in the OCC that showed significant activation in the initial localization experiment and was defined individually for each subject (see Levy et al., 2011, for details).

Posterior insula. The second cluster is in the posterior-middle portion of the left insula (which we refer to as pInsula). The same anatomical location was used for all subjects ($5 \times 5 \times 5$ mm cube, centred around Talairach coordinates -41, -16, 7), and was verified to be uncorrelated with the rank order of choices within individuals.[9]

In addition to the previous literature suggesting that both OCC and the pInsula are not related to valuation, these regions were selected based on reduced-form evidence that they will serve as useful corrections for measurement error (detailed in Section 3). These were the only grey matter regions we considered in our preliminary analysis. Neither of the regions were selected based on their performance in the model in Section 4, and they were the only regions examined with that model. For this reason, we will refer to them as *control* regions.

[9] The region we study here is anatomically separate from the portion of the anterior insula that, depending on the study, has been reported to correlate either positively or negatively with valuation (Bartra et al., 2013). For this reason, we conservatively chose voxels in the posterior portion of the insula (2–3 mm from the posterior tip).

Please cite this article as: R. Webb, N. Mehta and I. Levy, Assessing consumer demand with noisy neural measurements. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.07.028.

2.2. Data

The dataset we consider was collected in an experiment published by Levy et al. (2011). The purpose of the experiment was to assess whether neural activity measured during the passive viewing of choice objects could be used to predict subsequent choice behaviour. The experiment was thus divided into three stages, with the first two stages implemented inside an MRI scanner. In the first stage, each subject passively viewed the outcome of a series of small lotteries over changes to their wealth. The purpose of this stage was to identify the clusters of voxels which correlated with value. In the second stage, subjects passively viewed 20 consumer items while intermittently performing an incentivized task so as to maintain subject engagement. The purpose of this stage was to make repeated measurements for each item from the value areas identified in the first stage. Immediately after the second stage, subjects performed a third stage outside of the scanner in which they made all possible binary choices over this set of items in an incentive-compatible manner. Upon completion, each subject also received a $25 show-up fee in cash. We now describe the stages in detail.

2.2.1. Localization of subjective value in mPFC and vSTR

The first stage of the experiment was designed to independently identify the voxels which encode the subject's valuation of choice objects. Each subject was endowed with $40. On ensuing trials a lottery with equal probability of gaining or losing $2 was presented visually to the subject. The outcome of the lottery was then revealed to the subject and the result was added to or deducted from the subject's wealth. In total, 128 trials of this kind were presented. Within each subject, only the voxels which exhibited a statistically significant response to the outcome (winning or losing) were identified as our region of interest. Perhaps not surprisingly, voxels within both the mPFC and the vSTR were identified using this method, and constituted the regions of interest for the following stage. Levy et al. (2011) provide details of the regions of interest (ROIs), as well as example images.

2.2.2. Measuring the valuations of consumer items

Immediately following the first stage, subjects completed a second stage in the scanner intended to measure the subjective values of 20 consumer items. Subjects completed six 7-minute brain scans over the course of 45 min, each consisting of 40 trials, for a total of 240 trials. In each of these trials, subjects passively viewed an image of one of 20 different items, including four DVD movies, two books, four art posters, three music CDs, two pieces of stationery, and five monetary lotteries represented by pie charts. Each lottery offered a 50% chance of receiving a designated amount of money ($10, $15, $20, $25, $30) and a 50% chance of receiving $0. All items were presented 12 times in a random order to each subject. Subjects were instructed that "when they saw an item they should think about how much it was worth to them in a dollar amount".

To keep subjects alert, on 20 randomly selected trials (one for each of the 20 items), subjects were asked whether they preferred the item they had just seen or a randomly selected amount of money (ranging from $1 to $10). Subjects were told that one of these question trials would be randomly realized at the end and they would receive their selection on that trial (the item or the money). All subjects responded to these trials, and these 20 question trials were excluded from all behavioural and neural analysis. During the scanning stage, subjects did not know they would subsequently be offered an opportunity to choose between these same items after the scanning process was complete.

The measurement data is therefore a panel consisting of a sequence of neural measurements from $C = 11$ consumers.[10] The neural measurements consist of $M = 11$ BOLD measurements for each of $J = 20$ consumer items, randomly sequenced for a total of 220 measurements, per subject. We report measurements from $R = 4$ different regions (clusters of voxels), the mPFC, vSTR, the OCC as originally defined in the Levy et al. (2011) study, and the pInsula as defined in Section 2.1.[11]

2.2.3. Choice task

After a 5 min delay following the second scanning stage, subjects were asked to perform a choice task outside of the scanner. Subjects were presented with a complete series of binary choices between the 20 items previously presented in the scanner. Each possible binary comparison (190 choices) was presented twice (switching the left–right location on each repetition), in random order, for a total of 380 choices. The result of one of these choices was randomly selected for realization.

The choices of subjects were largely consistent, with $96 \pm 2\%$ of triplets transitive. Subjects switched their selection in $9 \pm 1\%$ of choice repetitions. Though there was a degree of homogeneity typical in an undergraduate subject pool (Fig. 1), choices were idiosyncratic across subjects. The individual preferences of a given subject could not be predicted simply by examining the preferences of other subjects (the mean correlation of the choice-ranking between pairs of subjects, excluding lotteries was 0.1 ± 0.3).

[10] Neural data from all regions was only available for 11 of the 12 participants originally reported in the dataset. As reported earlier, meta-analyses which aggregate over subjects have been conducted (Levy and Glimcher, 2012; Bartra et al., 2013; Clithero and Rangel, 2013).

[11] The BOLD signal was averaged across voxels within each cluster and over a time window of 4–6 TRs (a TR is the repetition time of a pulse cycle in the measurement technology, set to 1 TR = 2 s). This method for extracting the BOLD measurement from the pInsula are identical to those used for the mPFC, vSTR, and OCC in the original study. Readers are referred to Levy et al. (2011) for more technical details.

Please cite this article as: R. Webb, N. Mehta and I. Levy, Assessing consumer demand with noisy neural measurements. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.07.028.

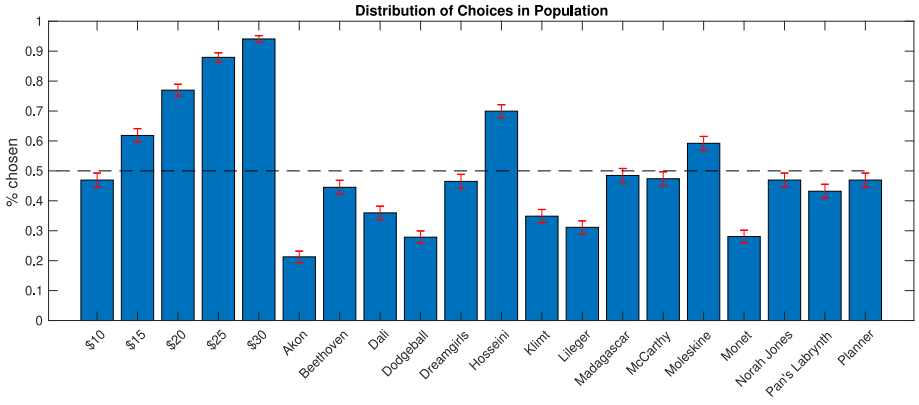

Fig. 1. Distribution of choices in the population.

Table 1
Estimates of logit model of choice on the difference in neural measurement between alternatives (standard errors in parentheses).

		Value Regions	Control Regions
mPFC	β_1	0.194	
		(0.073)	
vSTR	β_2	0.942	
		(0.125)	
OCC	β_3		0.084
			(0.063)
pInsula	β_4		−0.107
			(0.084)
LL ($n = 4180$)		−2840	−2896
BIC		5696	5809

3. Reduced form evidence

Our approach to alleviating the error-in-variables problem requires using contemporaneous signals from control regions as proxies for the measurement error in value regions. We now outline two conditions required for this approach and provide reduced form evidence in support of them.[12] Following that, we discuss a third condition, which although not directly related to the proxy variable approach, clarifies why a simple method to reduce the measurement error problem will not work in this context.

Condition 1. *The neural activity levels in control regions are unrelated to the value of the alternatives under consideration (i.e. they are not correlated with subsequent choices of those alternatives).*

This condition is otherwise known as the redundancy condition (Wooldridge, 2002), which states that if measurement error could be observed, or if there was no such measurement error in the first place, then the proxy variables must not explain any variation in utility. This condition thus requires that the neural activity in control regions of the brain is independent of valuation. To verify this condition in our dataset, we estimated a logit regression of consumer choice based on Eq. (3). The observable regressors were either the average BOLD signals associated with alternative j from the two value regions of the brain, mPFC and vSTR, or the two control regions of the brain, OCC and pInsula.[13] Table 1 reports the estimation results of these two models. As would be expected from earlier work (e.g. Knutson et al., 2007), the estimates for both the mPFC and vSTR in the first model are positive and highly significant, implying that both signals impact utility ($\chi^2 = 115$, $p < 0.000$). However, in the second model, the coefficients on OCC and pInsula measurements are not significant ($\chi^2 = 2.71$, $p < 0.258$). This suggests that both OCC and pInsula do not significantly influence utility, thus satisfy Condition 1.

The next condition is equally critical for the control regions to serve as proxies for the measurement error in mPFC and vSTR.

[12] Ultimately, assumptions underlying valid instruments and/or proxy variables are not fully testable, therefore always rely to some extent on *a priori* judgement. Though previous meta-analyses demonstrate the control regions have no correlation with valuation, our results are therefore still only dispositive.

[13] The randomized ordering of the binary choices in the experiment makes an intercept term redundant.

Please cite this article as: R. Webb, N. Mehta and I. Levy, Assessing consumer demand with noisy neural measurements. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.07.028.

Table 2
Correlation matrix for measurements from 4 brain regions.

	mPFC	vSTR	OCC
mPFC	–		
vSTR	0.512	–	
OCC	0.374	0.398	–
pInsula	0.304	0.306	0.213

Condition 2. *Contemporaneous measurement errors are significantly correlated across both valuation and control regions of the brain.*

Evidence for this condition can be found in the correlation matrix between contemporaneous measurements in the four regions we study (Table 2). Not surprisingly, activity in the mPFC and vSTR is highly correlated, likely due to their previously established role in the valuation of choice alternatives. However we also find that BOLD measurements in OCC and pInsula are significantly correlated both with each other and with the mPFC and vSTR. This suggests support for Condition 2. Given that activity in OCC and pInsula is not correlated with valuation (Table 1), it follows that the correlations between OCC, pInsula, and valuation regions must be due to some other factor, which for the purposes of this study, manifests as measurement error.

Finally, we consider whether a simple alternative approach can be used to reduce the measurement error problem. Recall that the neural measurements consist of $M = 11$ signals for each alternative and subject. The sample mean of the $M = 11$ measurements for a given alternative could then be used as a regressor in the choice model. If measurement error were substantially reduced by doing so, then there would be no need for our proposed method. The following condition rules out this approach.

Condition 3. *The measurement error is substantial even if we average the BOLD signal across the M measurement trials.*

To verify this condition, we examined the response in the signal in the two value regions of the brain (mPFC, vSTR) to the presentation of each alternative. In particular, we assessed how much variation in the BOLD signal in each of two value regions of a given subject could be explained by the presentation of the different alternatives to that subject during the measurement stage. The unexplained residual variation in the BOLD signal provides us with an estimate of the variance in the BOLD measurement due to measurement error (details of this analysis are given in the Appendix). We estimate that just over half of the overall variance in the sample mean (mPFC: 0.527; vSTR: 0.536) can be attributed to measurement error.

4. Model

4.1. Neural random utility model

We now describe a neural random utility model of consumer choice. This model is a reduced-form representation of the predominant computational decision models in the neuroscience literature (Fehr and Rangel, 2011; Webb, 2019). Each consumer $c \in \{1 \dots C\}$ is given a total of T pairwise choices amongst $j \in \{1 \dots J\}$ alternatives. We only specify the consumer index c when it is critical, so all variables should be taken to be consumer specific. For example, when presented with alternative j at choice occasion t, let U_{jt} be a consumer's utility for j. Similarly, let $n_{r,jt}$ be the activity level of value-encoding neurons in region r of the consumer's brain. The consumer's utility for an alternative is a linear combination of the contemporaneous neural activity levels in mPFC and vSTR ($r = 1$ and 2, respectively), governed by parameters β_1 and β_2.[14]

$$U_{jt} = \beta_1 n_{1,jt} + \beta_2 n_{2,jt}.$$

There are two issues inherent in bringing this model to a prediction dataset. First, U_{jt} is defined as a function of the contemporaneous neural activity levels, that is, during trial t of the choice task. However in our dataset, we observe neural measures for each alternative during the preceding measurement stage. It follows that the neural activity levels during a choice trial may vary from those measured during the measurement stage, simply due to the stochastic nature of neural activity. Second, we do not perfectly measure the activity of value-encoding neurons during the measurement stage. Instead, we measure the noisy BOLD signal from a pre-defined region of the brain that includes additional sources of

[14] This formulation is quite general. We only require that an aggregate statistic can represent the activity level of all value-encoding neurons in these regions. For example, $n_{1,jt}$ could be a (weighted) linear combination of the activity levels in each voxel in mPFC which encodes value. Similarly, it could be a linear combination of the spike rates of neurons within each of those voxels. Note that this formulation does not place any restriction on the sign or magnitude of the weights in this aggregation. We also do not require that *only* mPFC and vSTR contribute to utility. Value-related activity from other regions would either enter utility directly (if observed) or as an error term (if not), which would carry through in our calculations below. Of course, Condition 1 would still need to be assessed for any potential control region.

Please cite this article as: R. Webb, N. Mehta and I. Levy, Assessing consumer demand with noisy neural measurements. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.07.028.

variation. Together, these issues imply that we must transform utility in terms of the measured BOLD signal for alternative j during the measurement stage.[15]

We make this transformation in two steps. First, we separate the deterministic and stochastic component of neural activity,

$$n_{r,jt} = N_{r,j} + \nu_{r,jt},$$

where $N_{r,j} \equiv E_t[n_{r,jt}]$ is the expected neural activity for alternative j in a brain region (where the expectation is over trials within consumer and alternative) and $\nu_{r,jt}$ is the deviation from this expectation. Therefore the utility of alternative j at choice occasion t can be written in terms of the expected neural activity,

$$U_{jt} = \beta_1 N_{1,j} + \beta_2 N_{2,j} + \epsilon_{jt}, \tag{6}$$

where the error term $\epsilon_{jt} \equiv \beta_1 \nu_{1,jt} + \beta_2 \nu_{2,jt}$ represents the unobserved fluctuations in value associated with alternative j. We assume it to be IID across all consumers, all alternatives and all choice occasions.

Next, we replace the neural activity levels $N_{r,j}$ with the noisy BOLD signals from the measurement stage. Let $B_{r,jm}$ be the observed BOLD signal associated with alternative j in region r during measurement trial $m \in M$. We represent it as a linear function of the neural activity level on the measurement trial,

$$B_{r,jm} = \gamma(n_{r,jm} + \mu_{r,jm}),$$

where $\mu_{r,jm}$ is error in the measurement of value and γ is the unit scaling between BOLD and neural activity.[16] Note that, under this formulation, the term $\mu_{r,jm}$ captures both error in the measurement of the BOLD signal (e.g. for technical or physiological reasons) as well as error arising for functional reasons (e.g. neural activity in the pre-defined region of mPFC voxels due to some other function, perhaps in concert with other brain regions).[17] Substituting the expected value of the neural activity levels, $N_{r,j}$, yields

$$B_{r,jm} = \gamma(N_{r,j} + \nu_{r,jm} + \mu_{r,jm}). \tag{7}$$

The RHS of Eq. (7) consists of two terms, $\nu_{r,jm}$ and $\mu_{r,jm}$, which make up the error on a measurement trial. The former represents the deviation from expectation in the neural activity encoding value, and the latter represents the deviation in the BOLD signal from this neural activity. Therefore both of these errors vary over different consumers, alternatives, and measurement trials, but critically, the $\mu_{r,jm}$ may also be correlated with different regions of the brain on a measurement trial. Averaging Eq. (7) over the M measurement trials, we get the following expression for the expected neural activity level as a function of the sample means over measurements,

$$N_{r,j} = \frac{1}{\gamma}\bar{B}_{r,j} - \bar{\nu}_{r,j} - \bar{\mu}_{r,j}.$$

Substituting this expression for $N_{r,j}$ into Eq. (6) yields

$$U_{jt} = \alpha_1 \bar{B}_{1,j} + \alpha_2 \bar{B}_{2,j} + \xi_j + \epsilon_{jt}, \tag{8}$$

where

$$\xi_j \equiv -(\beta_1 \bar{\nu}_{1,j} + \beta_2 \bar{\nu}_{2,j} + \beta_1 \bar{\mu}_{1,j} + \beta_2 \bar{\mu}_{2,j}). \tag{9}$$

Eq. (8) specifies utility in terms of the average BOLD signals from mPFC and vSTR, measured for alternative j. The parameters of interest, $\alpha_1 \equiv \frac{\beta_1}{\gamma}$ and $\alpha_2 \equiv \frac{\beta_2}{\gamma}$, capture the marginal contribution of neural activity in mPFC and vSTR to consumer's utility (β_1 and β_2), scaled by the unit transformation of the BOLD signal γ. Since both the contribution to utility from each region and the scale of BOLD might differ over consumers, we will allow for unobserved heterogeneity in α_1 and α_2 over consumers.

The term ξ_j is the overall measurement error (referred to simply as the measurement error henceforth). It is composed of two terms from each of the value regions; the first ($\bar{\nu}_r$) arises because value measurements were not made concurrently with choice trials, and the second ($\bar{\mu}_r$) is the sample mean of measurement error in the signal.[18] The measurement error ξ_j is therefore IID over consumers and alternatives. However, unlike the ϵ_{jt}, it is invariant over repeated choice occasions for a given consumer and alternative, and may be correlated with measurements from other brain regions.

The impact of ξ_j on the parameter estimates is twofold. Not only will estimates of α_1 and α_2 be imprecise, but more importantly, they will be biased towards zero because the regressors $\bar{B}_{1,j}$ and $\bar{B}_{2,j}$ are negatively correlated with the measurement error ξ_j via Eqs. (7) and (9). We now discuss two different approaches for alleviating this problem.

[15] In the neuroscience literature, it is common to implement experimental designs in which choices are made concurrently with neural measurements. In such cases, we would only be concerned with the second issue (see footnote 18 for more detail).

[16] We can safely ignore an intercept term in this specification. Since it will be common across the measurements of all alternatives, ultimately, it will not impact the differences in measurements between alternatives.

[17] In addition, error in the pre-definition of the voxels comprising the mPFC region of interest will be captured by $\mu_{1,jm}$, as well as error in the pre-specified weighting (or sign) of those voxels.

[18] When measurements are made concurrently with choices, the measured activity on the current trial, $B_{r,jt}$, enters utility directly (rather than an average activity over measurement trials). Then $\nu_{r,jt}$ should no longer be considered measurement error. However, the remaining source, $\mu_{r,jt}$, may still be correlated over regions and our proposed correction for the error-in-variables problem would still be feasible.

Please cite this article as: R. Webb, N. Mehta and I. Levy, Assessing consumer demand with noisy neural measurements. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.07.028.

4.2. Alleviating the error-in-variables problem

4.2.1. Instruments for measurement error

We first consider whether additional measurements of the BOLD signal can be used as instruments for the endogenous regressors. In a standard IV approach to the errors-in-variables problem, the goal is to find an observable that is correlated with the regressors $B_{1,j}$ and $B_{2,j}$ in utility, but uncorrelated with the measurement error. Typically, additional measurements of the same latent variables are used as instruments. In the context of a neuroeconomic dataset, we can therefore consider whether additional measurements of the BOLD signal associated with the same alternative and consumer from the same regions (mPFC and vSTR) can be used as instruments. To serve as valid instruments, these additional measurements must be (i) exogenous, that is, they should be independent of the measurement errors in utility, and (ii) relevant, that is, they should be significantly correlated with the regressors, $B_{1,j}$ and $B_{2,j}$, in utility. Instruments which fail to satisfy these conditions can lead to sizeable finite sample bias in the estimates.

Recall that, in the measurement stage, there are $M = 11$ different measurements of the BOLD signal per alternative and consumer, with the alternative sequenced randomly over trials. Thus an IV approach can be implemented by specifying utility with the average of the first $K < M$ measurements associated with alternative j from mPFC and vSTR (instead of averaging over all M signals as in Eq. (8)) and taking the averages of the remaining $M - K$ measurements from that region *and* other value regions as instruments. These instruments satisfy the condition of exogeneity because in the measurement stage, there was a lengthy (and random) time gap between any two BOLD measurements of alternative j, implying that measurement errors in the first K signals will be independent of measurement errors in the following $M - K$ signals.[19] They might also be relevant because a measurement from one valuation region (e.g. mPFC) should correlate with a measurement from that same region on a different measurement trial — both are measurements of the same underlying valuation for an alternative, therefore they should be correlated. The same logic holds for a correlation in activity between valuation regions (e.g. mPFC and vSTR), even if these measurements are taken on different trials. This provides two potential instruments for two endogenous regressors.

A standard test for relevance is whether the F-statistic from a first-stage regression of the endogenous variables on the potential instruments is greater than a critical value of 10 (Staiger and Stock, 1997).[20] We therefore run the following first-stage regressions

$$\bar{B}^{K}_{1,j} = \kappa_1 + \theta_{11}\bar{B}^{M-K}_{1,j} + \theta_{12}\bar{B}^{M-K}_{2,j} + \nu_{1,j},$$

$$\bar{B}^{K}_{2,j} = \kappa_2 + \theta_{21}\bar{B}^{M-K}_{1,j} + \theta_{22}\bar{B}^{M-K}_{2,j} + \nu_{2,j},$$

where the sample means $\bar{B}^{K}_{1,j}$ and $\bar{B}^{K}_{2,j}$ are the regressors in utility, and $\bar{B}^{M-K}_{1,j}$ and $\bar{B}^{M-K}_{2,j}$ are the two instruments. While we do observe significant correlation between the split measurement samples, unfortunately, for all positive values of K, the largest value of the F-stat from either of these regressions was 6.503. This "weakness" in the instruments is to be expected if the measurement error is large enough such that splitting the measurement sample does not yield a strong enough correlation between the two halves. The poor fit implies that the instruments are too weak and a standard IV approach cannot be used to control for measurement error in this dataset.

Moreover, weak instruments will likely be an issue for any fMRI dataset. Even after averaging over K and M-K signals, the measurement error in the sample mean is high. While this problem can potentially be alleviated with a large number of independent measurements per alternative per consumer, the limited number of possible measurement trials in typical fMRI experimental designs (due to cost) makes this infeasible. Therefore the IV approach for dealing with the measurement error problem seems infeasible for neural datasets given current technology.

4.2.2. Proxies for measurement error

We now propose an alternative approach for alleviating the measurement error problem in fMRI data. Note that the measurement error ξ_j in Eq. (8) is simply an omitted variable that is correlated with the regressors $\bar{B}_{1,j}$ and $\bar{B}_{2,j}$. If we can find an observable that captures this correlation, then including that variable in the utility specification can alleviate the error-in-variables problem, provided it also satisfies the redundancy condition noted in Section 3. The effectiveness of a *proxy* variable thus depends on the extent to which it is able to capture the variation in the measurement error (Wooldridge, 2002). The larger the correlation between the proxy and measurement error, the greater will be the alleviation of the error-in-variables problem.

The proxy variables we consider are the sample means of the M BOLD measurements from two control regions of the brain (the OCC and the pInsula), where each measurement was taken contemporaneously when the consumer was exposed to alternative j. We index these control regions by $r = \{3, 4\}$ and denote the proxies by $\bar{B}_{r,j} \equiv \frac{1}{M}\sum_{m=1}^{M} B_{r,jm}$. As described in Section 2, these proxies appear to satisfy the criterion of redundancy. First, the control regions were chosen

[19] Section 2 discusses correlations in the measurement error between contemporaneous signals from different brain areas. However, the exogeneity condition requires independence in the measurement error *across different measurement trials*, which we would expect to be zero.

[20] On average, the size of the finite sample bias of an IV estimator relative to that of the OLS estimator is approximately equal to 1/F - the inverse F statistic of the first-stage IV regression in which the endogenous variable(s) are regressed on the instruments. Therefore the larger the value of the F statistic, the smaller will be the finite sample bias of estimates of α_1 and α_2.

Please cite this article as: R. Webb, N. Mehta and I. Levy, Assessing consumer demand with noisy neural measurements. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.07.028.

a priori based on a large body of evidence that they do not play a role in valuation. Second, neither of the BOLD measures from the control regions were significantly correlated with choice behaviour (Table 1). What remains to test is whether, and to what extent, these control variables are correlated with the measurement error in the BOLD signal across regions of the brain, thus can serve as useful proxy variables.

To see how these proxy variables enter our utility specification, first define the measurement error as a linear function of the two proxy variables,

$$\xi_j = \alpha_3 \bar{B}_{3,j} + \alpha_4 \bar{B}_{4,j} + \varsigma_j,$$

where ς_j is the residual measurement error (IID over consumers and alternatives), and the parameters α_3 and α_3 determine the extent to which the proxy variables capture the variation in the measurement error ξ_j.[21] Substituting this expression into Eq. (8) yields,

$$U_{jt} = \alpha_1 \bar{B}_{1,j} + \alpha_2 \bar{B}_{2,j} + \alpha_3 \bar{B}_{3,j} + \alpha_4 \bar{B}_{4,j} + \varsigma_j + \epsilon_{jt}, \tag{10}$$

where $\alpha_1 = \frac{\beta_1}{\gamma}$ and $\alpha_2 = \frac{\beta_2}{\gamma}$ are still the parameters of interest.

The inclusion of proxy variables can only reduce the asymptotic bias in the estimates of α_1 and α_2 compared to the baseline model (8), not eliminate it (Wickens, 1972; McCallum, 1972). This reduction depends on the extent to which the proxy variables capture the variation in the measurement error. If the proxies are positively correlated with the measurement error, then we should expect the estimates of α_3 and α_4 to both be less than zero. To see why, consider the case when a measurement error for an alternative is large and positive. Then the BOLD signal in control regions will be large (due to measurement error), but so will the measurements in value regions. In fact, these value measurements will be *too large*, therefore their contribution to utility should be attenuated on this trial. Therefore if the proxy variables are working as intended, we should expect an increase in the estimates of α_1 and α_2 and negative estimates for α_3 and α_4.[22]

It also remains to verify whether the inclusion of the proxy variables will result in more or less precise estimates of α_1 and α_2 and a smaller estimate for the variance of the measurement error, ξ_j. In principle, the standard errors of the estimates and the estimated variance of the measurement error could either increase or decrease after including proxy variables (Aigner, 1974). The intuition is as follows. We can think of proxy variable(s) as having two components: (i) the common measurement error, which is the measurement error that is common between the regressor and the proxy variable, and (ii) variation in the proxy variable that is uncorrelated with both the measurement error and the regressor. When we add the proxy variable to the utility, a portion of the previously unobserved measurement error is now explained by component (i) of the proxy variable. This results in a reduction of the variance of measurement error in the utility. However, adding the proxy variable can also increase the variance because of the addition of component (ii).[23] Thus there are two opposing forces at work and the net change in the variance of measurement error will depend on their relative magnitudes.

4.3. Estimation details

To illustrate the ability of proxy variables to address the error-in-variables problem, we estimate five nested models. All proposed models are estimated using simulated maximum likelihood. Technical details of the estimation procedure are in Appendix B.

Model 1: Logit. The first is a simple Logit model, based on the utility in Eq. (8), which ignores measurement error. The temporal fluctuation in value, ϵ_{jt}, is assumed to be IID Extreme Value over choice occasions, consumers, and alternatives, with variance normalized to one.[24] This model consists of 2 parameters: (i) α_1 and α_2, the relationship between signals in the mPFC and vSTR and utility. Results for this model were already reported in Table 1.

Model 2: Baseline random-effect. The second model uses a random-effect to model the alternative-specific measurement error, $\xi_j \sim \mathcal{N}(0, \sigma^2_{ME})$. The scale of the measurement error σ_{ME} is identified because ξ_j is IID over consumers and alternatives, but is invariant over all choice occasions for a given consumer and alternative. This model consists of three parameters, α_1, α_2, and σ_{ME} the scale of the measurement error

Model 3: Proxy variable correction. The third model is our proposed approach to alleviate the measurement error problem via a proxy variable, based on the utility given in Eq. (10). This specification adds the coefficients of the two proxy variables (α_3 and α_4) to Model 2, for a total of five parameters.

---

21 We have removed the intercept term since it is common across the utilities of all alternatives.

22 More formally, recall from Eq. (9) that the sources of measurement error ξ_j enter with a negative sign. Next recall from Table 2 that the signals $\bar{B}_3$, $\bar{B}_4$ are positively correlated with the signals $\bar{B}_1$, $\bar{B}_2$. Now since the coefficient of the measurement error in the utility is positive (unity), it implies that α_3 and α_4 will be negative.

23 Note that component (ii) is uncorrelated with the regressors, thus it does not bias coefficient estimates.

24 The IID EV distribution is given by $F(\epsilon) = \exp(-\exp(-\epsilon))$. This distribution implies a logit expressions for the choice probabilities.

Please cite this article as: R. Webb, N. Mehta and I. Levy, Assessing consumer demand with noisy neural measurements. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.07.028.

Table 3
Parameter estimates and model comparison of Logit, Random-effect, and Proxy variable correction.

		1: Logit		2: Baseline random-effect		3: Proxy variable correction	
		Estimate	(S.E.)	Estimate	(S.E.)	Estimate	(S.E.)
mPFC	α_1	0.1935	(0.0731)	0.2796	(0.1025)	0.4046	(0.1049)
vSTR	α_2	0.9424	(0.1258)	1.9148	(0.1834)	2.3901	(0.1985)
OCC	α_3					−0.1350	(0.0672)
pInsula	α_4					−1.8701	(0.1313)
Meas. Err.	σ_{ME}			1.4516	(0.0536)	1.4566	(0.0486)
LL (N = 4180)		−2840.1		−1951.1		−1880.8	
BIC (N = 4180)		5696.9		3927.2		3803.2	

Models 4-5: Heterogeneity in utility coefficients. The fourth and fifth models augment the specification by allowing for unobserved heterogeneity across consumers in the specification of utility. Specifically, the parameters α_1 and α_2 in (8) are assumed to be independently normally distributed across consumers with means $\tilde{\alpha}_1, \tilde{\alpha}_2$ and variances σ_1^2, σ_2^2. Therefore the relationship between the neural measurement and utility can vary across subjects when σ_1^2 or σ_2^2 are non-zero. The fourth model consists of five parameters: (i) $\tilde{\alpha}_1$ and $\tilde{\alpha}_2$, the population level means of the relationship between signals in the mPFC and vSTR on consumer preference, (ii) σ_1 and σ_2, the standard deviations of this relationship across consumers; and (iii) σ_{ME}, the scale of the measurement error. The fifth adds two additional proxy variables, α_3 and α_4, for a total of seven parameters.

5. Empirical results

In Section 5.1, we begin with results from the simpler specifications without consumer heterogeneity, as they clearly demonstrate the ability of the proxy variables to account for measurement error. In Section 5.2, we present the results from incorporating consumer heterogeneity. All of these results are presented for the full sample to reduce the variance of our estimates. In Section 5.3 we consider a specification with alternative-specific indicators, and in Section 5.4, we consider an out-of-sample validation exercise.

5.1. Model estimates without heterogeneity in utility coefficients

Results for the model without heterogeneity are presented in Table 3. In the Random-Effect specification, the estimates of both α_1 (mPFC) and α_2 (vSTR) are positive, significant, and larger than in the simple Logit model. This result is consistent with previous studies finding that an increase in mPFC and vSTR activity increases the probability of choosing an item (e.g. Knutson et al., 2007; Webb et al., 2019). However when the proxy variables for measurement error are included together with mPFC and vSTR, the estimates of both α_1 and α_2 increase substantially. This suggests that the Logit estimates are significantly biased when measurement error is not controlled for.

Moreover, both α_3 and α_4 enter significantly negative, as expected. Recall that the coefficients for the control regions, by themselves, were not significant choice predictors (Table 1). Since they are positively correlated with the measurement error — therefore with the signals from mPFC and vSTR — they serve to attenuate the measurement of utility when included in the utility specification. We also observe an improvement in fit from the proxy variable correction and a drop in the BIC.

The standard deviation of the measurement error remains consistent at 1.45 times than that of ϵ. The standard errors of all estimates remain consistent across specifications. As discussed in Section 4.2.2, including proxy variables leads to two countervailing forces that increase and decrease the variance of the measurement error in the utility. In this dataset, these two forces appear to balance.

5.2. Model estimates with heterogeneity in utility coefficients

Including proxy variables to correct for measurement error has similar effects when we allow for heterogeneity in the BOLD response across consumers (Table 4). In particular, the means of the random coefficients both increase after controlling for measurement error and are consistent with the estimates of α_1 and α_2 in Table 3. Again, this suggests a substantial reduction in bias. Our estimate of the standard deviation of the measurement error is 1.31 (Model 4) and 1.59 (Model 5) times that of the fluctuations in valuation ϵ_{jt}.

The results from the heterogeneous model also highlight an interesting implication of measurement error: controlling for measurement error yields a substantial decrease in the variance of the random coefficients. This suggests that a significant degree of the heterogeneous BOLD responses typically observed across subjects in such studies may be due to the degree of measurement error within consumer. In particular, once measurement error is controlled for, the variance of the vSTR coefficient (σ_2) is not significantly different than zero. To interpret this result, it is instructive to consider the sources of heterogeneity in α_1 and α_2.

Please cite this article as: R. Webb, N. Mehta and I. Levy, Assessing consumer demand with noisy neural measurements. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.07.028.

Table 4
Model estimates allowing for heterogeneity across consumers in BOLD response.

		4: Baseline random-effect		5: Proxy variable correction	
		Estimate	(S.E.)	Estimate	(S.E.)
mPFC	$\tilde{\alpha}_1$	0.0149	(0.1017)	0.8928	(0.1015)
vSTR	$\tilde{\alpha}_2$	1.3421	(0.1995)	3.6333	(0.1998)
OCC	α_3			−1.0393	(0.0948)
pInsula	α_4			−0.9824	(0.1518)
Meas. Err.	σ_{ME}	1.3091	(0.0454)	1.5849	(0.0509)
Std. Dev. of mPFC	σ_1	0.7707	(0.1180)	0.2965	(0.1056)
Std. Dev. of vSTR	σ_2	0.2321	(0.1659)	0.0111	(0.1751)
LL (N = 4180)		−1948.3		−1874.8	
BIC (N = 4180)		3938.3		3759.6	

First, heterogeneity might arise because the scale (γ) between neural activity and the BOLD measure might differ over consumers. This is a common interpretation of the random effect in fMRI modelling. Second, the contribution to utility from the mPFC and/or vSTR (β_1, β_2) might differ over consumers. However if σ_2 is indeed small once the measurement error on each trial is controlled for, then that suggests that neither source of heterogeneity are present across measurements from the vSTR in this sample. This finding might explain why signals from the vSTR have been so effective in predicting consumer choices across consumers and field settings (Venkatraman et al., 2015; Genevsky et al., 2017): the relationships between neural activity, the BOLD signal, and the choice probabilities appear to be stable across consumers.

By contrast, there is still variation in the coefficient for the mPFC (σ_1) even after controlling for measurement error. Given that our estimate of σ_2 suggests that there is little variation in γ across consumer in this sample, we can attribute the remaining variation in the relationship between mPFC and choice as arising in the contribution to utility β_2. Again, this is consistent with the hypothesis that the role of mPFC in valuation might be more idiosyncratic across consumers (Knutson and Genevsky, 2018).

5.3. Alternative-specific indicators

In typical consumer choice studies, alternative-specific indicators are often included so that the baseline choice probability for an alternative is isolated. Therefore in our study, the coefficients on alternative-specific dummies represent the average valuation (over consumers) for each of the 20 items. When they are included in the model, the coefficients on the valuation signals are identified purely from the deviations of the individual consumers signals from the mean individual.

In our results presented so far, we have examined models without intercepts for two reasons. First, an objective of the neuro-prediction literature is to assess the degree to which latent neural signals in the value regions can capture utility without pre-specifying the identity of the alternative (e.g. Smith et al., 2014). Second, the estimates of a model with alternative-specific dummies cannot be used to predict the market shares/choices of products that were not included in the estimation sample. Such would be the case when evaluating new products, or in the case of our application in Section 5.4, when predicting across product categories. If the purpose of a prediction exercise is to get an estimate of the baseline choice probability for an alternative, a model with alternative-specific indicator assumes it can be estimated directly from choice data.

However it is still useful to consider a model with alternative-specific indicators. First, we can assess whether the signals from value regions contain information beyond the simple identity of the alternative. This is particularly important because our full sample contains 5 lotteries with $-amounts that increase monotonically. We should ensure that our results are not simply driven by a monotonic relationship in these lotteries. Second, because the alternative-specific indicators control for the average valuation in the sample, the coefficients on the value signals will then measure the degree to which value signals capture deviations in consumer valuation from the sample average. This will allow us to test hypotheses about the role of different brain regions.

We report the estimates with alternative-specific indicators in Table 5. The role of the proxy variables under this specification is consistent with the results previously reported. The estimates of both α_1 and α_2 are still positive and significant, and increase substantially with the proxy variable correction. Our estimate of the standard deviation of the measurement error is consistent at 1.3 times that of the fluctuations in valuation ϵ_{jt}. The significance of the alternative-level indicator variables is evidence that there is considerable correlation in the preferences of NYU undergraduates in the Levy et al. (2011) sample. There is also a dramatic improvement in the fit. This suggests that there is a substantial amount of variance in the choice data that is not accounted for solely by the value signals, nor completely corrected for by our proxies for measurement error.

The degree to which the coefficients change when including alternative-specific indicators is also interesting (compared to Model 3, see Fig. 2). When the sample-average valuations are controlled for, the coefficient for vSTR decreases by over 50% (2.3901 vs. 1.3844, $p < 0.001$). By contrast, the decrease in the coefficient for mPFC across specifications is smaller

Please cite this article as: R. Webb, N. Mehta and I. Levy, Assessing consumer demand with noisy neural measurements. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.07.028.

Table 5
Model estimates including brand dummy variables.

		4: Baseline random-effect		5: Proxy variable correction	
		Estimate	Std Err	Estimate	Std Err.
mPFC	α_1	0.0551	(0.1169)	0.2239	(0.1182)
vSTR	α_2	0.9038	(0.2055)	1.3844	(0.2150)
OCC	α_3			−0.1009	(0.1174)
pInsula	α_4			−1.5452	(0.1487)
Meas. Err.	σ_{ME}	1.6559	(0.0587)	1.7326	(0.0613)
Akon	λ_1	–		–	
Beethoven	λ_2	1.4602	(0.2098)	1.2050	(0.208)
Dali	λ_3	1.5472	(0.2194)	0.9969	(0.2198)
Dodgeball	λ_4	1.4196	(0.2110)	0.5883	(0.2081)
Dreamgirls	λ_5	2.5933	(0.2107)	2.7464	(0.2153)
Hosseini	λ_6	4.8232	(0.2468)	4.9970	(0.2519)
Klimt	λ_7	1.2759	(0.2143)	1.0890	(0.2187)
Lileger	λ_8	2.4091	(0.2149)	2.1094	(0.2144)
Madagascar	λ_9	1.9815	(0.2118)	2.0866	(0.2142)
McCarthy	λ_{10}	1.7940	(0.2059)	1.5345	(0.2046)
Moleskine	λ_{11}	2.7467	(0.2201)	2.7579	(0.2220)
Monet	λ_{12}	1.0870	(0.2139)	0.3408	(0.2103)
Norah Jones	λ_{13}	2.5358	(0.2236)	2.0210	(0.2190)
Pan's Labrynth	λ_{14}	2.2545	(0.2157)	2.0767	(0.2187)
Planner	λ_{15}	3.2640	(0.2196)	2.7237	(0.2142)
$10 Lottery	λ_{16}	1.8747	(0.2071)	1.0056	(0.2082)
$15 Lottery	λ_{17}	4.6623	(0.2438)	4.2870	(0.2388)
$20 Lottery	λ_{18}	4.5473	(0.2461)	4.2053	(0.2405)
$25 Lottery	λ_{19}	5.6134	(0.2756)	5.1147	(0.2708)
$30 Lottery	λ_{20}	6.7798	(0.3188)	6.9765	(0.3239)
LL (N = 4180)		−1501.2		−1474.2	
BIC (N = 4180)		3208.3		3140.2	

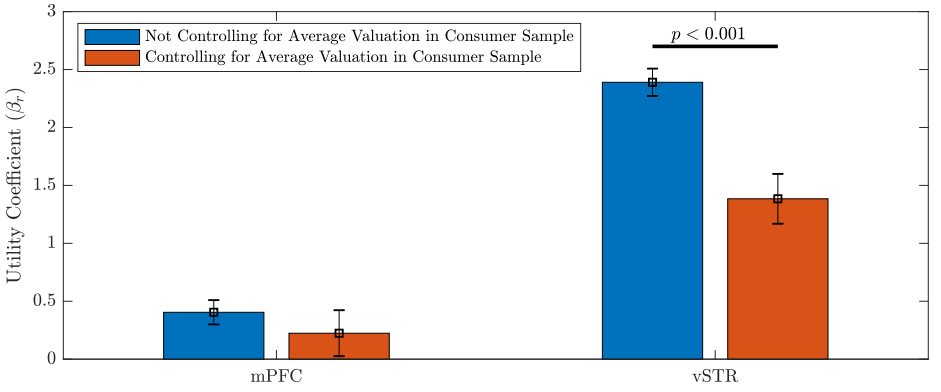

Fig. 2. The change in utility coefficients for different value regions (mPFC and vSTR) after controlling for average valuations across consumers.

and not significantly different (0.4046 vs. 0.2239, $p = 0.223$). The relative changes of the utility coefficients under this specification is consistent with the hypothesis that activity in the vSTR might represent a more common component of utility that is shared across consumers, perhaps because it codes for a more affective response (Knutson and Genevsky, 2018).

5.4. Predicting demand

A proposed advantage of neuroeconomic methods is a richer datasource on which to assess consumer demand. This includes estimating the marginal effect of manipulating a characteristic, either in isolation, or in conjunction with, traditional observables (Telpaz et al., 2015; Boksem and Smidts, 2015; Venkatraman et al., 2015; Genevsky et al., 2017). The presence of error in neural measurements, however, will bias such estimated marginal effects, with important implications for demand prediction.

To quantify the degree of this bias in our dataset and assess predictive performance, we conduct two validation exercises in which we split the overall sample into an estimation sample and a hold-out sample. The estimation sample

Table 6
Model estimates on sub-sample with no lotteries.

		1: Logit		2: Baseline random-effect		3: Proxy variable correction	
		Estimate	(S.E.)	Estimate	(S.E.)	Estimate	(S.E.)
mPFC	α_1	0.4442	(0.0731)	1.0166	(0.1482)	1.1741	(0.1556)
vSTR	α_2	0.0094	(0.1696)	0.7217	(0.2663)	2.5645	(0.3494)
OCC	α_3					−0.8104	(0.1481)
pInsula	α_4					−0.5559	(0.1767)
Meas. Err.	σ_{ME}			2.0134	(0.0867)	2.0795	(0.0817)
LL (N = 2310)		−1586.0		−973.7		−954.3	
BIC (N = 2310)		3187.6		1970.7		1947.2	

Table 7
Prediction results on sample of lotteries (N = 1870).

	1: Logit	2: Baseline random-effect	3: Proxy variable correction
Log likelihood	−1282.4	−746.5	−730.2
Pseudo R^2	0.011	0.424	0.437
MSPE	460.8	451.9	438.2

consists of the binary choices between the fifteen non-lottery items made by all 11 consumers (for a total of 320 × 11 = 2310 observations, since each consumer had to repeat each binary choice pair). We will define the hold-out for each exercise shortly.

We first re-estimate three models on the sub-sample of data that does not contain the lotteries: the simple Logit model (Model 1), the baseline random-effect model (Model 2), and our proxy variable correction (Model 3). Results are reported in Table 6, and are consistent with those reported in Sections 5.1 and 5.3. After including proxy variables to correct for measurement error, we observe a positive and significant estimates for α_1 and α_2 and negative estimates for α_3 and α_4. To assess the predictive performance of our approach, we calculated fit metrics for a hold-out sample consisting of all remaining choices between the five lotteries as well as each pair consisting of a lottery and a consumer good (Table 7). The out-of-sample Log Likelihoods, Pseudo R^2, and Mean-Squared Prediction Error all decrease after the measurement error correction.[25][26]

One feature of the lottery hold-out sample provides an ideal prediction test-case for assessing the degree of bias in the sample. Because the dollar amounts of these lotteries increased monotonically ($10, $15, $20, $25, and $30 if win, and $0 if lose), it is possible to analyse the change in demand as the lottery amount increases, relative to the reference $10 lottery. Unlike the other alternatives in our dataset, subjects with completely transitive preferences should always choose the higher lottery (relative to the $10 reference lottery). Indeed, this is what we find in our data; all consumers chose the relatively higher-valued lottery. This allows the lotteries to be ordered in a manner that is homogeneous across all consumers (unlike the other goods for which consumers had more idiosyncratic preferences), and implies that the model which predicts a higher choice probability for a higher-valued lottery will be a better model.

Therefore to assess the degree of bias, we also computed the predicted probabilities for each lottery using the difference between the average BOLD activity and the $10 lottery. Since the proxy areas are solely used to control for measurement error, thus achieve less-biased estimates, we calculated the predicted probabilities only using the estimated coefficients for the mPFC and vSTR from Table 7. These fitted probabilities are reported in Fig. 3. We find that our corrected estimates yields a predicted change in demand that is over twice as large as our baseline model (relative to chance), and over four times larger than a basic Logit specification.

6. Conclusion

We have proposed a method to directly control for the error-in-variables problem inherent in relating choice behaviour to noisy neural measurements. This issue has remained unexamined so far in the literature, but it has severe implications for estimates of the contribution to a utility specification from different brain regions. Our method considerably reduces the bias in these estimates, providing improved inference and performance in an out-of-sample demand prediction exercise. Perhaps most importantly, our method does not place stringent properties on the experimental design, such as repeated choice trials or additional calibration trials. It is thus widely applicable, both to the large number of existing datasets in the literature and to future studies.

[25] In this exercise, the out-of-sample likelihood is calculated by integrating over the distributions of the unobservables (measurement error and the control signals). Since the joint likelihood for a given individual consists of the product of the likelihood across all choices, it is crucial to account for this correlation across choices when computing the out-of-sample likelihood.

[26] In discrete choice applications, the pseudo R^2 typically range from 0.20 to 0.55, with the best-fitting models including alternate specific dummies and information on product characteristics (which we have not included). At the subject level, the proxy variable correction improves prediction results for 8 of the 11 consumers, has no effect for one, and performs worse for 2 of the 11.

Please cite this article as: R. Webb, N. Mehta and I. Levy, Assessing consumer demand with noisy neural measurements. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.07.028.

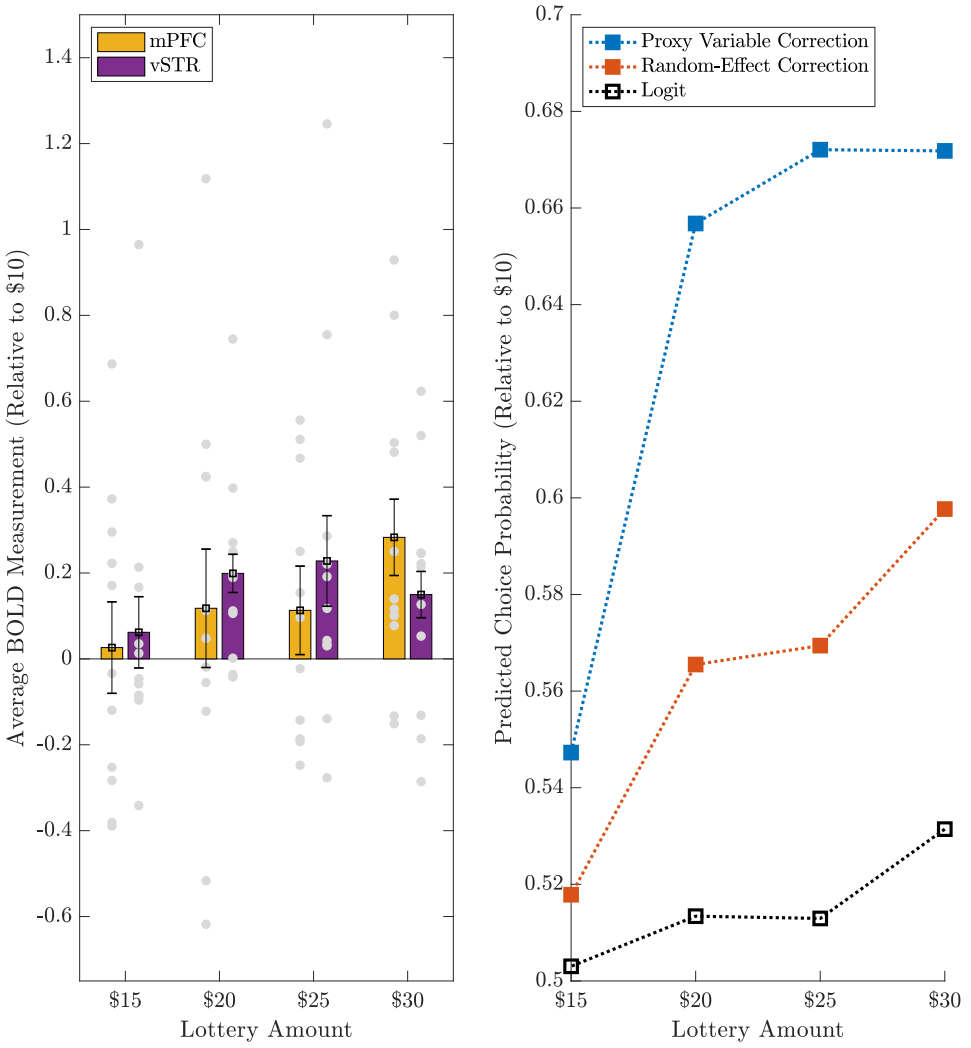

Fig. 3. Out-of-sample prediction of lottery choices. (Left) The average BOLD measurements for each lottery amount, relative to the $10 lottery. The average activity for each subject is also plotted in grey. (Right) The predicted choice probabilities based on model estimates, at the average BOLD measurements for each lottery. The actual choice probabilities in the sample are 1, since the lotteries are monotonic.

Our proposed method relies on the econometrician's choice of a proxy variable that is uncorrelated with choice behaviour but still correlated with signals from value regions of the brain. We considered two candidate control areas, the OCC and the pInsula, and our empirical results are consistent with a substantial reduction in the bias caused by the error-in-variables problem. We chose these two control areas based on previous literature and verified their suitability via a simple reduced-form analysis, but the tests that we have conducted so far should be seen as dispositive and not conclusive. Indeed, there are also many more possible brain regions we could have analysed. We did not conduct a full search for proxy variables that might satisfy the required proxy variable conditions. In principle, a reduced-form first-stage could be automated to identify "optimal" control variables amongst the large number of voxels we did not study, leading to a greater reduction in the bias of the estimated marginal effects.

A reduction in the bias of marginal effects is desirable for multiple reasons. Not only does it provide more accurate demand predictions based on neural measurements, but will also sharpen inference about the role of different brain areas in contributing towards utility. After we control for measurement error, we find a substantial decrease in the heterogeneity in the marginal coefficients of utility. For the vSTR in particular, we find that the variation in its contribution to utility is near zero across consumers in our sample. This suggests that the scale between neural activity and the choice probabilities might be stable across consumers in this brain region, at least under controlled conditions. Further work is needed to

Please cite this article as: R. Webb, N. Mehta and I. Levy, Assessing consumer demand with noisy neural measurements. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.07.028.

determine if this result generalizes across subject samples, and perhaps more importantly, across experimental designs which manipulate the distribution of choice objects that consumers might encounter.

A reduction in bias also will considerably reduce the occurrence of Type II errors in neuroeconomic studies. While the problem of "false negatives" is generally not considered as serious as "false positives" (and rightly so), a bias in estimated marginal effects towards zero actually make the conclusions of recent studies conservative; the true effect is likely larger than reported. Moreover, when the relationship between a neural variable and choice is biased towards zero, this will increase the probability that a false null hypothesis about the role of a brain area might fail to be rejected. In other words, using our proposed measurement error correction can help clarify whether studies which *do not currently* find a significant role for mPFC or vSTR in predicting behaviour do so because of measurement error, or because these regions are indeed not-predictive in that context. It also helps clarify why adding signals from additional brain regions like the Insula might sharpen inference about the role of mPFC and vSTR (e.g. Knutson et al., 2007). In particular, the analyst must exercise caution when inferring whether control regions are indeed involved in valuation, per se, or simply capturing the effects of measurement error.

Our study also sounds a cautionary note about the interpretation of results from modern machine-learning classifiers in neuro-prediction studies. Since the goal of such classifiers is solely prediction, they search the voxel space for the combination of signals which maximize predictive accuracy via a trade-off between bias and variance. Our results suggest that such methods will include voxels from regions of the brain that reduce the bias of the prediction model due to correlation in measurement error. Of course, for the purposes of prediction, this is completely reasonable. However interpreting the resulting output of these classifiers as "valuation maps" is clearly fraught with issues. As we demonstrate, signals from some regions of the brain can improve predictive performance even if they are not directly involved in valuation. Future research should consider how to interpret the voxel maps from such studies, and how they might inform the choice of proxy variables in our proposed method.

One of the sources of measurement error in a neuro-prediction exercise arises undoubtedly from variation in attention during measurement trials. The scanning portion of the experiment did contain an attention check — subjects were asked to choose between an item on the screen and an amount of money. All subjects registered a choice on these trials. Of course, this attention could waver over trials. Future studies should consider possible techniques for assessing attention, including eye-tracking analysis, to be used as an additional control.

Finally, we believe our proxy-variable method can be broadly applied. We have considered the BOLD signal from fMRI scanning because it is a widely-used measure of neural activity in the neuroeconomics literature. However in applied research, much of the market research performed in the commercial sector use cheaper, less precise methods such as EEG. Our proposed methodology is equivalently applicable to such measurement technologies, provided that suitable proxy variables can be found.

Appendix A. Relative degree of measurement error

To assess the degree of measurement error in the BOLD signal in each value region, we compare an estimate of the unexplained variance of the BOLD measurement after presentation of an alternative to the overall variance of the BOLD signal. This initial estimate is achieved by regressing the BOLD signal measured during the measurement stage (in each of the two value regions, for each subject) on the presentation of the different alternatives to that subject. The residual variation in the BOLD signal in this regression will provide us with an estimate of the variance in the BOLD measurement due to measurement error.

Since each consumer is presented with J alternatives, there are a total of $M \times J$ measurement trials per consumer, with the M measurements randomly dispersed throughout. We estimate a model of the $M \times J$ neural measurements across all consumers $c \in C$. Let $S_{c,r,m}$ be the BOLD measurement in brain region $r \in A$ when a consumer is exposed to an alternative during measurement trial m. We represent it as a function of the presentation of alternatives and a stochastic component

$$S_{c,r,m} = \alpha_{1,c} + \mathbf{I}_m^j \boldsymbol{\alpha}_2 + \mathbf{I}_m^{c,j} \boldsymbol{\alpha}_3 + \epsilon_{c,r,m} \tag{11}$$

where $\mathbf{I}^j$ is a vector of dummy variables for the presentation of item j across all consumers, and $\mathbf{I}^{c,j}$ is for the presentation of alternative j within each consumer. The J-vector $\boldsymbol{\alpha}_2$ therefore represents the shift in measured neural activity when all consumers are exposed to alternative j (relative to alternative 1), and the $(J-1)(C-1)$ parameter vector $\boldsymbol{\alpha}_3$ represents the consumer-specific shift. The error term, ϵ, is assumed to be independent over measurement trials, consumers, and brain areas.

Evidence for Condition 3 can be found in the residuals of Eq. (11). The error term in this regression is akin to the measurement error in the signal: it is the variation in BOLD activity not explained by the presentation of the alternative. OLS provides an estimate of the variance of this error, $\hat{\sigma}_r^2$, for each brain region. The estimate of the variance from the sample mean of the measurement error is therefore $\frac{\hat{\sigma}_r^2}{M}$. We are interested in the relative size of this variance relative to the overall variance of the sample mean $\frac{1}{M}\sum_{m=1}^{M} S_{c,r,m}$ over all alternatives. By definition it is smaller, though we still find that $\frac{\hat{\sigma}_r^2}{M}$ comprises just over half of the overall variance (mPFC: 0.527; vSTR: 0.536). We can safely say that the measurement error is substantial even if we average the BOLD signal across the M measurement trials.

Please cite this article as: R. Webb, N. Mehta and I. Levy, Assessing consumer demand with noisy neural measurements. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.07.028.

Appendix B. Estimation procedure

We state here the estimation procedure for the model pooled across all subjects (i.e. without unobserved heterogeneity). The parameters can be represented by a vector $\Theta \equiv [\theta_1, \sigma]$, where $\theta \equiv [\alpha_1, \alpha_2, \alpha_3, \alpha_4]$. Recall that all consumers were given T pairwise choices amongst the J alternatives in the choice stage. Consider the case where consumer c was given a choice amongst alternatives $j, k \in J$ during choice occasion t. The probability of choosing alternative j over k, conditional on the measurement errors for the two alternatives, $\varsigma_{c,j}$ and $\varsigma_{c,k}$, is

$$\Pr\left(d_c^{j,k} = 1|\theta, \varsigma_{c,j}, \varsigma_{c,k}\right) = \text{Logit}\left(\sum_{a=1}^{A} \alpha_a \left(\bar{B}_{a,c,j} - \bar{B}_{a,c,k}\right) + \varsigma_{c,j} - \varsigma_{c,k}\right),$$

where $d_c^{j,k} = 1$ is a binary indicator if alternative j is chosen over k. Consumer c's likelihood over all her pairwise choices conditional on the consumer specific measurement error draws $\{\varsigma_{c,j}\}_{j=1}^{J}$ is therefore

$$L_c\left(\theta, \{\varsigma_{c,j}\}_{j=1}^{J}\right) = \prod_{k=1}^{J-1}\prod_{k>j}^{J}\left[\Pr\left(d_c^{j,k} = 1|\theta, \varsigma_{c,j}, \varsigma_{c,k}\right)^{d_c^{j,k}} \times \Pr\left(d_c^{j,k} = 0|\theta, \varsigma_{c,j}, \varsigma_{c,k}\right)^{1-d_c^{j,k}}\right].$$

We next simulate $R = 30{,}000$ Halton draws of each of the errors in $\{\varsigma_{c,j}\}_{j=1}^{J}$ from $N(0, \sigma^2)$ to get the unconditional likelihood of all pairwise choices as

$$\hat{L}_c\left(\theta, \sigma_{ME}\right) = \frac{1}{R}\sum_{r=1}^{R} L_c\left(\theta, \{\varsigma_{c,j}\}_{j=1}^{J}\right).$$

Given the unconditional likelihood for consumer c, the sample log likelihood across all consumers will be

$$l(\Theta) = \sum_{c=1}^{C} \log \hat{L}_c(\theta, \sigma_{ME}),$$

and the SML estimates are given by

$$\Theta_{SML=} \arg\max_{\theta} l(\Theta).$$

References

Aigner, D.J., 1974. MSE dominance of least squares with errors-of-observation. J. Econometrics 2 (4), 365–372.
Barron, H.C., Dolan, R.J., Behrens, T.E.J., 2013. Online evaluation of novel choices by simultaneous representation of multiple memories. Nature Neurosci. 16 (10), 1492–1498.
Bartra, O., McGuire, J.T., Kable, J.W., 2013. The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. NeuroImage 76, 412–427.
Boksem, M.A.S., Smidts, A., 2015. Brain responses to Movie Trailers Predict Individual Preferences for Movies and their population-wide commercial success. J. Mark. Res. 52 (4), 482–492.
Clithero, J.A., Rangel, A., 2013. Informatic parcellation of the network involved in the computation of subjective value. Soc. Cogn. Affect. Neurosci. 9 (9), 1–14.
Delgado, M.R., Beer, J.S., Fellows, L.K., Huettel, S.A., Platt, M.L., Quirk, G.J., Schiller, D., 2016. Viewpoints: Dialogues on the functional role of the ventromedial prefrontal cortex. Nature Neurosci. 19 (12), 1545–1552.
Falk, E.B., Berkman, E.T., Lieberman, M.D., 2012. From Neural Responses to Population Behavior: Neural Focus Group predicts population-Level Media Effects. Psychol. Sci. 23 (5), 439–445.
Fehr, E., Rangel, A., 2011. Neuroeconomic Foundations of Economic Choice–Recent Advances. J. Econ. Perspect. 25 (4), 3–30.
Genevsky, A., Knutson, B., 2015. Neural affective Mechanisms Predict Market-Level Microlending. Psychol. Sci. 26 (9), 0956797615588467–1422.
Genevsky, A., Yoon, C., Knutson, B., 2017. When brain beats behavior: Neuroforecasting Crowdfunding Outcomes. J. Neurosci. 37 (36), 8625–8634.
Gratton, C., Laumann, T.O., Nielsen, A.N., Greene, D.J., Gordon, E.M., Gilmore, A.W., Nelson, S.M., Coalson, R.S., Snyder, A.Z., Schlaggar, B.L., Dosenbach, N.U.F., Petersen, S.E., 2018. Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation. Neuron 98 (2), 439–452.e5.
Gross, J., Woelbert, E., Zimmermann, J., Okamoto-Barth, S., Riedl, A., Goebel, R., 2014. Value signals in the prefrontal cortex predict individual preferences across reward categories. J. Neurosci. 34 (22), 7580–7586.
Hare, T.A., Camerer, C.F., Knoepfle, D.T., Rangel, A., 2010. Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. J. Neurosci. 30 (2), 583–590.
Harris, K.M., Keane, M.P., 1999. A model of health plan choice:: Inferring preferences and perceptions from a combination of revealed preference and attitudinal data. J. Econometrics 89 (1–2), 131–157.
Hsu, M., Krajbich, I., Zhao, C., Camerer, C.F., 2009. Neural response to reward anticipation under risk is nonlinear in probabilities. J. Neurosci. 29 (7), 2231–2237.
Karmarkar, U.R., Shiv, B., Knutson, B., 2015. Cost conscious? The neural and behavioral impact of Price Primacy on Decision Making. J. Mark. Res. 52 (4), 467–481.
Khaw, M.W., Grab, D.A., Livermore, M.A., Vossler, C.A., Glimcher, P.W., 2015. The measurement of subjective value and its relation to contingent valuation and environmental Public Goods. PLoS One 10 (7), e0132842.
Knutson, B., Genevsky, A., 2018. Neuroforecasting aggregate choice. Curr. Dir. Psychol. Sci. 27 (2), 110–115.
Knutson, B., Rick, S., Wimmer, G.E., Prelec, D., Loewenstein, G., 2007. Neural predictors of purchases. Neuron 53 (1), 147–156.

Please cite this article as: R. Webb, N. Mehta and I. Levy, Assessing consumer demand with noisy neural measurements. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.07.028.

Lebreton, M., Jorge, S., Michel, V., Thirion, B., Pessiglione, M., 2009. An automatic valuation system in the human brain: evidence from functional neuroimaging. Neuron 64 (3), 431–439.
Levy, D., Glimcher, P.W., 2011. Comparing apples and oranges: Using reward-specific and reward-general subjective value representation in the brain. J. Neurosci. 31 (41), 14693–14707.
Levy, D.J., Glimcher, P.W., 2012. The root of all value: a neural common currency for choice. Curr. Opin. Neurobiol. 22 (6), 1027–1038.
Levy, I., Lazzaro, S., Rutledge, R.B., Glimcher, P.W., 2011. Choice from non-choice: predicting consumer preferences from blood oxygenation level-dependent signals obtained during passive viewing. J. Neurosci. 31 (1), 118–125.
Logothetis, N.K., 2002. The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 357 (1424), 1003–1037.
Lusk, J.L., Crespi, J.M., McFadden, B.R., Cherry, J.B.C., Martin, L., Bruce, A., 2016. Neural antecedents of a random utility model. J. Econ. Behav. Organ. 132, 93–103.
McCallum, B.T., 1972. Relative Asymptotic Bias from Errors of Omissions and measurement. Econometrica 40 (4), 757–758.
McClure, S.M., Li, J., Tomlin, D., Cypert, K.S., Montague, L.M., Montague, P.R., 2004. Neural correlates of behavioral preference for Culturally Familiar Drinks. Neuron 44 (2), 379–387.
McFadden, D.L., 2001. Economic choices. Amer. Econ. Rev. 91 (3), 351–378.
McFadden, D.L., 2013. The new science of pleasure. In: Hess, S., Daly, A. (Eds.), Handbook of Choice Modelling. books.google.com.
Plassmann, H., O'Doherty, J.P., Rangel, A., 2007. Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. J. Neurosci. 27 (37), 9984–9988.
Plassmann, H., O'Doherty, J.P., Shiv, B., Rangel, A., 2008. Marketing actions can modulate neural representations of experienced pleasantness. Proc. Natl. Acad. Sci. 105 (3), 1050–1054.
Smith, A., Bernheim, B.D., Camerer, C.F., Rangel, A., 2014. Neural activity Reveals Preferences without Choices. Am. Econ. J. Microecon. 6 (2), 1–36.
Staiger, D., Stock, J.H., 1997. Instrumental variables regression with Weak instruments. Econometrica 65 (3), 557.
Telpaz, A., Webb, R., Levy, D.J., 2015. Using EEG to predict Consumers' Future Choices. J. Mark. Res. 52 (4), 511–529.
Tusche, A., Bode, S., Haynes, J.-D., 2010. Neural responses to unattended products predict later Consumer Choices. J. Neurosci. 30 (23), 8024–8031.
Venkatraman, V., Dimoka, A., Pavlou, P.A., Vo, K., Hampton, W., Bollinger, B., Hershfield, H.E., Ishihara, M., Winer, R.S., 2015. Predicting advertising success beyond traditional measures: New insights from neurophysiological methods and market response modeling. J. Mark. Res. 52, 436–452.
Webb, R., 2019. The (Neural) dynamics of Stochastic Choice. Manage. Sci. 65 (1), 230–255.
Webb, R., Levy, I., Lazzaro, S., Rutledge, R.B., Glimcher, P.W., 2019. Neural random utility: Relating cardinal neural observables to stochastic choice behaviour. J. Neurosci. Psychol. Econ. 12 (1), 45–72.
Wickens, M.R., 1972. A note on the use of Proxy Variables. Econometrica 40 (4), 759–761.
Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. The MIT Press.
Yatchew, A., Griliches, Z., 1985. Specification Error in Probit Models. Rev. Econ. Stat. 67 (1), 134–139.

Please cite this article as: R. Webb, N. Mehta and I. Levy, Assessing consumer demand with noisy neural measurements. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.07.028.