Online Appendices

Appendix A: The ellipsoidal design

This appendix presents in more detail the ellipsoidal design, which was summarised in Section 3. As described there, for each question $k \in \{1, \ldots, K\}$, the algorithm has two steps.

(1) The selection of question k is guided by the criterion of minimising "volumetric D-error", which is calculated given a multivariate normal distribution of partworths $\beta \sim \mathcal{N}(\mu, \Sigma)$, where $\mu = \mu^{k-1}$ and $\Sigma = \Sigma^{k-1}$.

(2) Given an answer to question k, the mean and variance are updated from μ^{k-1}, Σ^{k-1} to μ^k, Σ^k, in accordance with Bayes's rule.

These two steps are described in the next two sections; the second step is described first. For more details on the ellipsoidal design, see SV.

A1. Updating the mean and variance of the partworth distribution

How, in response to an answer to question k, are the mean and covariance are updated from μ^{k-1} and Σ^{k-1} to μ^k and Σ^k? Let question k be denoted by the set $\{x, y\}$ with $x \neq y$ (interpreted as a choice between x and y). The prior distribution of β is assumed to be normally distributed $\mathcal{N}(\mu, \Sigma)$ with $\mu = \mu^{k-1}$ and $\Sigma = \Sigma^{k-1}$. Let $\Sigma^{\frac{1}{2}}$ denote the square root of Σ according to the Cholesky decomposition $\Sigma = \Sigma^{\frac{1}{2}}(\Sigma^{\frac{1}{2}})^T$. Define the mean and standard deviation of question k as $\mu_{x,y} := (x - y) \cdot \mu$ and $\sigma_{x,y} := ||(\Sigma^{\frac{1}{2}})^T(x - y)||$. Suppose that x is chosen over y, so $U_x \geq U_y$. Proposition 1 in SV shows that, given the new data point, the mean and variance of the posterior distribution are:

$$\mathbb{E}(\beta | U_x \geq U_y) = \mu + (\Sigma^{\frac{1}{2}} W)_1 \mathbb{E}(Z) \tag{1}$$

$$Cov(\beta|U_x \geq U_y) = \Sigma^{\frac{1}{2}} W \begin{bmatrix} Var(Z) & (0_{J-1})^T \\ 0_{J-1} & I_{J-1} \end{bmatrix} (\Sigma^{\frac{1}{2}} W)^T \tag{2}$$

where

- W is a $J \times J$ matrix such that $W = [w_1, \ldots, w_J]$, where $\{w_j\}_{j=1}^J$ is an orthonormal basis such that $w_1 = (\Sigma^{\frac{1}{2}})^T (x-y)/\sigma_{x,y}$ and $(\Sigma^{\frac{1}{2}} W)_1$ is the first column of $\Sigma^{\frac{1}{2}} W$;
- Z is a random variable with density $f_z(z) = C^{-1}(1 + \exp(-\mu_{x,y} - \sigma_{x,y} z))^{-1} \phi(z)$ where $C = \int_{\mathbb{R}} (1 + \exp(-\mu_{x,y} - z\sigma_{x,y}))^{-1} \phi(z) dz$;
- 0_{J-1} is a vector of zeros of dimension $J-1$ and I_{J-1} an identity matrix of order $J-1$.

Thus the updated mean of the distribution is $\mu^k = \mathbb{E}(\beta|U_x \geq U_y)$ given by the expression in equation (1) and the updated variance is $\Sigma^k = Cov(\beta|U_x \geq U_y)$ in equation (2). The moment-matching approximation assumes that the posterior distribution is normally distributed with this mean and variance.

A2. Question selection using volumetric D-error

The ellipsoidal design employs a variant of the widely used efficiency criterion "D-error". SV call their variant "volumetric D-error". To specify this criterion formally, consider any question $\{x, y\}, x \neq y$ and define $a_{x,y}$ to be a random variable that takes on a value of 1 if x is chosen (i.e. if $U_x \geq U_y$) and 2 otherwise. The volumetric D-error for question $\{x, y\}$ is:

$$\mathbb{E}_{a_{x,y}} \Big(\det(\mathrm{Cov}(\beta|a_{x,y}))^{1/d} \Big) \tag{3}$$

where $d = 2$. The standard specification of D-error sets d in equation (3) equal to the number of parameters, J, but SV motivate their variant of D-error by pointing to its relationship to a "natural credibility region" (SV, p.320). Ideally, the selection of questions should be determined by the policy which solves the dynamic programming problem of minimizing the volumetric D-error after all K question have been asked. This problem is intractable, however. So the ellipsoidal method solves a one-step-ahead problem. That is, given the posterior distribution after $k-1$ questions, the k^{th} question is chosen to minimise (3). To solve the problem of minimising the error in (3), SV define the function $g(m, v)$ as follows:

$$g(m, v) := p(m, v)\mathrm{Var}(Z(m, v))^{1/d} + p(-m, v)\mathrm{Var}(Z(-m, v))^{1/d} \tag{4}$$

where $Z(m,v)$ is a random variable with density $f_{Z(m,v)}(z) = p(m,v)^{-1}(1+\exp(-m-vz))^{-1}\phi(z)$; and $p(m,v) = \int_{\mathbb{R}}(1+\exp(-m-zv))^{-1}\phi(z)dz$. Suppose that, prior to question k, the distribution of β had a covariance matrix $\Sigma = \Sigma^{k-1}$. Proposition 2 in SV shows that the volumetric D-error from question $\{x,y\}$ is:

$$\mathbb{E}_{a_{x,y}}\Big(\det(\mathrm{Cov}(\beta|a_{x,y}))^{1/2}\Big) = \det(\Sigma)^{1/2}g(\mu_{x,y},\sigma_{x,y}) \tag{5}$$

According to the ellipsoidal method, the aim is to choose, as the k^{th} question, the question $\{x,y\}$ which minimises equation (5). But clearly this is proportional to the value of the function g. So the aim is to choose the question $\{x,y\}$ that minimises g. Like SV, we minimise an approximation to g, $\breve{g}$, which is obtained by calculating the values of g on a grid and thereby constructing a piecewise linear function that approximates g. Unlike SV, however, we find the question that exactly minimises $\breve{g}$, whereas SV use mixed integer programming (MIP) to find a near-optimal solution. Our approach is computationally feasible only if the number of attributes J is not too large; MIP is desirable when J is larger. Appendix C shows that we are able, broadly speaking, to replicate key results of SV for HB estimation regarding the metrics that we employ. Our replication results suggest that the performance of estimators is similar when MIP is used and when the exact optimal solution is found.

Appendix B: Estimation procedure

As noted in Section 5, like SV, we obtain HB estimates using the probabilistic programming language STAN, via its Julia interface. STAN estimates posterior distributions using Hamiltonian MCMC simulations. (See Gelman et al. (2013, Ch. 12)). Our assessment of the convergence of MCMC chains is guided by the analysis of diagnostics in Vehtari et al. (2021); most notably, their recommendation that it should be the case that (1) the scale reduction statistic $\hat{R} < 1.01$ and (2) the effective sample size statistic $ESS \geq 400$. In their HB estimation procedure, SV make a simplifying assumption, which has substantial implications for the convergence properties of the MCMC chains. They assume that the researcher knows the true value of the population parameter Σ; so they only estimate the population parameter μ and the individual parameters β^i. Given our focus on the population parameters, we do not make this simplifying assumption; in reality, the researcher typically will not know the

true value of the population covariance matrix. Thus we obtain estimates of the population parameters μ and Σ and the individual parameters. When we relax SV's assumption, however, the convergence properties of the MCMC chains are substantially worse. Moreover, they are much worse under the ellipsoidal design than the random design. SV (p.330) report that, under their simplifying assumption, they achieve convergence by running "four MCMC chains for 5,000 iterations plus the default 1,000 warmups". In contrast, when we estimate Σ as well as μ and β^i, we do not achieve convergence if we use the same number of questions, subjects and MCMC chain iterations. Convergence is achieved when we (1) increase the MCMC chain iterations to 25,000 for each of the four chains (2) increase the warm-ups per chain to 5,000, (3) increase the number of questions to 30 (4) increase the number of subjects to 150, and (5) assume that the covariance matrix Σ is diagonal (we estimate the $J = 10$ diagonal elements). These convergence problems are a particular problem for the ellipsoidal design; for datasets generated by the random design, the MCMC chains converge far more readily.

Appendix C: Replication of results in SV

Three metrics calculated by SV and us are (1) the normalised RMSE (denoted in Table 3 by $\overline{\text{RMSE}}$ β) (2) the hit-rate and (3) the market share MAE. For these three metrics, this appendix assesses our ability to replicate SV's results for data generated by the ellipsoidal design when the HB estimator is used. While SV obtain their results from a single simulated dataset, our results are obtained by simulating 50 datasets, and taking the average across simulations. Table 3 presents our results, which are broadly consistent with those of SV. For the RMSEs and the market share MAEs, our replicated estimates are all within 2.5 standard deviations of those of SV. Our estimates of the hit-rate, however, are somewhat higher than SV's estimates, even though we use the formula for the hit-rate in SV's code. The discrepancy between our estimates and SV's estimates of the hit-rate might arise from differences between the code SV supply and the code generating the results reported in SV's published article.[1] Our estimates of the hit-rate, however, display a similar dependency on

[1] We found a discrepancy between the code SV supply and the description in SV's footnote 4 regarding how they calculate the market share MAE. In our replications of the market share MAE, we follow the description

the number of questions to those of SV.

Table 3: Replication Results for the Ellipsoidal Design and HB Estimator†

	$\overline{\text{RMSE}}$ β			Hit-rate			Market share MAE		
# Questions	4	8	16	4	8	16	4	8	16
Low A, Low H.									
Replication	0.822	0.772	0.687	0.773	0.791	0.819	0.049	0.048	0.046
	(0.030)	(.028)	(.020)	(0.009)	(0.008)	(0.005)	(0.003)	(0.003)	(0.007)
SV	0.84	0.76	0.67	0.69	0.70	0.72	0.05	0.05	0.04
Low A, High H.									
Replication	0.510	0.456	0.373	0.859	0.874	0.899	0.054	0.043	0.033
	(0.015)	(0.013)	(0.008)	(0.007)	(0.006)	(0.004)	(0.004)	(0.003)	(0.004)
SV	0.50	0.45	0.37	0.81	0.82	0.84	0.06	0.05	0.03
High A, Low H.									
Replication	1.129	0.986	0.762	0.701	0.745	0.808	0.063	0.052	0.045
	(0.033)	(0.026)	(0.023)	(0.010)	(0.007)	(0.004)	(0.007)	(0.006)	(0.004)
SV	1.10	0.98	0.76	0.68	0.71	0.76	0.07	0.05	0.04
High A, High H.									
Replication	0.867	0.746	0.553	0.766	0.806	0.861	0.096	0.065	0.038
	(0.023)	(0.023)	(0.022)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.004)
SV	0.89	0.77	0.55	0.74	0.78	0.84	0.10	0.07	0.04

† Figures in brackets are standard deviations of replication estimates over the 50 simulations.

Appendix D: The effect of normalising the RMSE

We consider two metrics of estimation accuracy for the individual parameters: a normalised RMSE; and the unnormalised RMSE. The normalisation is from SV: in the normalised RMSE of the estimator, both it and the true values of the parameters are normalised so that the sum of the absolute values of the coefficients is equal to J. Toubia et al. (2004) and Toubia et al. (2007) also use a normalised RMSE, in order to facilitate comparison between scenarios. Toubia et al. (2004) suggest that the normalisation is innocuous: "this scaling does not change the relative comparison among question-design methods". We find, however, that this is not the case. In particular, we find that, while the normalised RMSE of β for the

in footnote 4. This allows us to successfully replicate their results for market share MAEs, which we cannot do using the code SV supply. SV's code is available at `https://github.com/juan-pablo-vielma/EMFACBCA`.

ellipsoidal design is lower than the random design in all four scenarios, this is reversed for the unnormalised RMSE in the two low heterogeneity scenarios. The remainder of this appendix explains how the normalisation might distort the assessment of estimation accuracy.

Consider two estimators, one which is biased and one which is not. If the biased estimator does not have a lower variance than the unbiased estimator, it will have a higher MSE, because the MSE is the sum of the variance and the square of the bias. But this advantage of the unbiased estimator may be masked if the estimators are normalised. We illustrate this with a simple example. Consider J independent random variables $X_j, j = 1, \ldots, J$ each with a mean θ_j. Suppose for each random variable X_j, there are D draws, $X_j^d, d = 1, \ldots, D$. Consider two estimators of θ_j. The first is the sample mean $\bar{\theta}_j = \frac{1}{D}\sum_{d=1}^{D} X_j^d$, which is an unbiased estimator. The second is $\hat{\theta}_j = \bar{\theta}_j(1 + B)$ for some $B > 0$. $\hat{\theta}_j$ is a biased estimator, with a bias of $E[\hat{\theta}_j - \theta_j] = B\theta_j$. Suppose we normalise $\hat{\theta}_j$ and $\bar{\theta}_j$ using the same type of normalisation as SV.

$$\hat{\theta}_j^* := \frac{J\hat{\theta}_j}{\sum_{i=1}^{J} |\hat{\theta}_i|} \quad , \quad \bar{\theta}_j^* := \frac{J\bar{\theta}_j}{\sum_{i=1}^{J} |\bar{\theta}_i|} \quad , \quad j = 1, \ldots, J$$

where the normalised estimators are estimators of the normalised true values of the parameters: $\frac{J\theta_j}{\sum_{i=1}^{J} |\theta_i|}$. But clearly the two normalised estimators are identical:

$$\hat{\theta}_j^* := \frac{J\hat{\theta}_j}{\sum_{i=1}^{J} |\hat{\theta}_i|} = \frac{J\bar{\theta}_j(1 + B)}{\sum_{i=1}^{J} |\bar{\theta}_i(1 + B)|} = \frac{J\bar{\theta}_j}{\sum_{i=1}^{J} |\bar{\theta}_i|} = \bar{\theta}_j^* \quad , \quad j = 1, \ldots, J$$

So the RMSEs of the two normalised estimators will be identical. Thus the normalisation masks a source of imprecision in the estimator $\hat{\theta}_j$ that arises from its bias.

Appendix E: Estimation accuracy and bias in larger datasets

Table 4 compares estimates of the bias for the ellipsoidal and random designs, not only for the base case but also for two variants with larger datasets: the first variant has 60 questions rather than 30; the second has 300 subjects rather than 150. For each of the three cases and four scenarios, the absolute value of the bias for the ellipsoidal design is substantially more than for the random design, suggesting the adaptive design generates endogeneity bias.

But, as suggested by Liu et al. (2007), in sufficiently large datasets, the endogeneity bias

Table 4: Estimator bias and the size of the dataset

	Bias β			Bias Σ			Bias μ		
# Subjects	150	150	300	150	150	300	150	150	300
# Questions	30	60	30	30	60	30	30	60	30
Low A, Low H.									
Ellipsoidal	0.158	0.052	0.078	0.250	0.073	0.124	0.156	0.050	0.078
Random	0.010	0.000	0.005	0.002	0.004	0.000	0.010	-0.001	0.006
Low A, High H.									
Ellipsoidal	0.130	0.033	0.057	0.316	0.084	0.139	0.130	0.035	0.060
Random	0.008	0.006	0.004	0.030	0.014	0.016	0.008	0.011	0.005
High A, Low H.									
Ellipsoidal	0.283	0.058	0.138	0.232	0.050	0.116	0.278	0.056	0.134
Random	0.021	0.018	0.026	0.021	0.018	0.012	0.020	0.019	0.032
High A, High H.									
Ellipsoidal	0.333	0.086	0.140	0.443	0.115	0.184	0.343	0.086	0.139
Random	0.087	0.024	0.037	0.127	0.053	0.053	0.079	0.025	0.034

generated by adaptive designs may be less problematic. Table 4 shows that, in each of the two variants of the base case, for each of the four scenarios, for each type of parameter, the absolute value of the endogeneity bias under the ellipsoidal design is lower than in the base case, because the datasets are larger. We need to investigate, therefore, whether, for the larger datasets, the reduction in endogeneity bias might reverse the finding in Table 1 that the ellipsoidal method generates higher RMSEs than the random method for the population parameters. Table 5 shows, however, that no such reversal takes place. For the population parameters, even in the larger datasets, the estimation accuracy is worse under the ellipsoidal design than under the random design.

Table 5: Estimation accuracy and the size of the dataset

	RMSE β			RMSE Σ			RMSE μ		
# Subjects	150	150	300	150	150	300	150	150	300
# Questions	30	60	30	30	60	30	30	60	30
Low A, Low H.									
Ellipsoidal	0.4875	0.3226	0.3909	0.2950	0.1093	0.1589	0.1965	0.0875	0.1073
Random	0.4122	0.3463	0.4061	0.0988	0.0586	0.0627	0.0725	0.0560	0.0487
Low A, High H.									
Ellipsoidal	0.6227	0.4307	0.5335	0.3617	0.1437	0.1898	0.1942	0.1095	0.1034
Random	0.6529	0.4989	0.6451	0.1234	0.0899	0.0853	0.1182	0.0938	0.0767
High A, Low H.									
Ellipsoidal	0.6770	0.4413	0.5812	0.2849	0.1042	0.1659	0.3472	0.1215	0.1965
Random	0.6707	0.5530	0.6628	0.1449	0.0916	0.0916	0.1268	0.1088	0.0949
High A, High H.									
Ellipsoidal	0.9636	0.6366	0.8018	0.5214	0.2123	0.2534	0.4364	0.2149	0.2239
Random	1.0560	0.7817	1.0401	0.2362	0.1516	0.1490	0.2099	0.1584	0.1499

References

Gelman, A., J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin (2013). *Bayesian Data Analysis, 3rd ed.* Boca Raton: Chapman and Hall, CRC.

Liu, Q., T. Otter, and G. Allenby (2007). Investigating endogeneity bias in marketing. *Marketing Science 26(5)*, 642–650.

Toubia, O., J. Hauser, and R. Garcia (2007). Probabilistic polyhedral methods for adaptive choice-based conjoint analysis: Theory and application. *Marketing Science 26(5)*, 596–610.

Toubia, O., J. Hauser, and D. Simester (2004). Polyhedral methods for adaptive choice-based conjoint analysis. *Journal of Marketing Research 61*, 116–131.

Vehtari, A., A. Gelman, D. Simpson, B. Carpenter, and P.-C. Bürkner (2021). Rank-normalization, folding and localization: An improved $\hat{R}$ for assessing convergence of MCMC (with discussion). *Bayesian Analysis 16(2)*, 667–718.