Web Appendix

Patient- or Physician-Oriented Marketing: What Drives Primary Demand for Prescription Drugs?

Marc Fischer and Sönke Albers

In this appendix, we provide details on model estimation, results, and robustness checks to check the appropriateness of the assumptions made for model specification.

MODEL ESTIMATION PROCEDURE

Recall that some of the unknown parameters in our brand sales model are specified as category- and brand-specific while others are not (see Equation 6 in the paper). To estimate this model we have data available for 2,831 brands over a period of 21 quarters starting in quarter 4/2000 until quarter 4/2005. Since the brand sales model includes lagged sales the first quarter is inherently lost. In addition, we note that many brands are launched later than quarter 4/2000 and a few brands are withdrawn from the market during the observation period. As a result, we have an unbalanced panel data set amounting to eventually 47,308 observations which we can use for estimation. Hence, we have on average 16.7 quarters per brand and 32.9 brands per category (86 categories) available.

Our brand sales model incorporates brand-specific effects which reflect brand heterogeneity in the market. Considering only own and competitive marketing expenditure variables, a fixed-effects specification would require estimating ~8,500 own brand sales effects and ~25,000 cross-effects associated with detailing, professional journal advertising, and DTC.[1] With 47,308 observations at hand, we need to impose some structure on the heterogeneous

[1] 2,831 brands × 3 own expenditure variables yield 8,493 own sales effects. For competitive expenditure variables, we have 2,831 brands × 3 competitive expenditure variables – 86 categories × 2 (1st entrant and 2nd entrant spending do not apply to pioneer + second entrant per category) = 8,321. Summing across three expenditure types yields 24,963 cross-effects.

parameters to keep the estimation feasible. Specifically, we assume the following heterogeneity structure (consistent with Equation 7 in the paper):

$$\theta_{lki} = \overline{\theta}_l + \lambda_{l1}\upsilon_{1ki} + \lambda_{l2}\upsilon_{2ki}, \text{ where } \upsilon_{1ki}, \upsilon_{2ki} \sim N(0,1) \text{ and } \mathrm{Cov}(\upsilon_{1ki}, \upsilon_{2ki}) = 0\,. \qquad \text{(W.1)}$$

where θ_{lki} represents an unknown brand-specific parameter associated with predictor l, $\overline{\theta}_l, \lambda_{l1},$ and λ_{l2} are heterogeneity parameters to be estimated, υ_{1ki} and υ_{2ki} denote variance components that vary by brand and category, and L is the number of predictors with heterogeneous parameters (l = 1, 2, …, L). The implied variance of θ_{lki} is ($\lambda_{l1}^2 + \lambda_{l2}^2$). The variance-covariance matrix for $\boldsymbol{\theta}_{ki}$ is given by $\boldsymbol{\Sigma} = \boldsymbol{\Lambda\Lambda'}$. Specifically,

$$\boldsymbol{\Sigma} = \begin{bmatrix} \lambda_{11}^2 + \lambda_{12}^2 & \lambda_{11}\lambda_{21} + \lambda_{12}\lambda_{22} & \cdots & \lambda_{11}\lambda_{L1} + \lambda_{12}\lambda_{L2} \\ \lambda_{21}\lambda_{11} + \lambda_{22}\lambda_{12} & \lambda_{21}^2 + \lambda_{22}^2 & \cdots & \lambda_{21}\lambda_{L1} + \lambda_{22}\lambda_{L2} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{L1}\lambda_{11} + \lambda_{L2}\lambda_{12} & \lambda_{L1}\lambda_{21} + \lambda_{L2}\lambda_{22} & \cdots & \lambda_{L1}^2 + \lambda_{L2}^2 \end{bmatrix}. \qquad \text{(W.2)}$$

Note that the structure of $\boldsymbol{\Sigma}$ is completely determined by the λ-heterogeneity parameters which are to be estimated. For each random parameter associated with predictor l, we estimate the mean, $\overline{\theta}_l$, and the two λ-heterogeneity parameters, λ_{l1} and λ_{l2}. Together with the non-random parameters, our estimation method needs to recover 157 unknown parameters.[2]

Let $\mathbf{y}_{ki}$ and $\mathbf{x}_{ki}$ denote time-series vectors that stack all observations on the dependent variable, i.e. log of sales, and the predictor variables for brand i in category k. Let $\boldsymbol{\Theta}$ and $\boldsymbol{\varphi}$ denote vectors of unknown parameters, covering both random and non-random parameters, and $\boldsymbol{\upsilon}_{ki}$ denote a standard normally distributed latent variable, which we need for the estimation of the variance of (heterogeneous) parameter distributions and the derivation of individual parameters. Then, the likelihood function for the brand sales model can be written out as (consistent with Equation 8 in the paper):

[2] This set of parameters includes 21 means and 42 λ-heterogeneity parameters for estimating the location and variance of random parameter distributions. In addition, we have 85 parameters for 86 categories, 3 parameters for 4 quarters and 6 parameters for non-random control variables (5 quality classes, order of entry, and availability of a radical innovation).

$$L = \prod_{k=1}^{K} \prod_{i=1}^{N_k - 1} \int_{\boldsymbol{\upsilon}_{ki}} f\left(\boldsymbol{\upsilon}_{ki}\right) f\left(\mathbf{y}_{ki} \middle| \mathbf{x}_{ki}, \boldsymbol{\Theta}, \boldsymbol{\varphi}, \boldsymbol{\upsilon}_{ki}\right) d\boldsymbol{\upsilon}_{ki}, \quad \text{with } \boldsymbol{\Theta} = \left(\bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}}, \bar{\boldsymbol{\gamma}}, \bar{\delta}, \boldsymbol{\lambda}\right),$$
$$\text{where } f\left(\mathbf{y}_{ki} \middle| \mathbf{x}_{ki}, \boldsymbol{\Theta}, \boldsymbol{\varphi}, \boldsymbol{\upsilon}_{ki}\right) = \prod_{t=1}^{T_i} f\left(\mathbf{y}_{kit} \middle| \mathbf{x}_{kit}, \boldsymbol{\Theta}, \boldsymbol{\varphi}, \boldsymbol{\upsilon}_{ki}\right). \tag{W.3}$$

The term $f\left(\mathbf{y}_{ki} \middle| \mathbf{x}_{ki}, \boldsymbol{\Theta}, \boldsymbol{\varphi}, \boldsymbol{\upsilon}_{ki}\right)$ denotes the marginal density of $\mathbf{y}_{ki}$. We estimate (W.3) by using the maximum likelihood technique. To account for heteroscedasticity, we use the market share based weight, w_{ki}, and maximize the following log likelihood function:

$$\log L = \sum_{k=1}^{K} \sum_{i=1}^{N_k - 1} w_{ki} \int_{\boldsymbol{\upsilon}_{ki}} f\left(\boldsymbol{\upsilon}_{ki}\right) f\left(\mathbf{y}_{ki} \middle| \mathbf{x}_{ki}, \boldsymbol{\Theta}, \boldsymbol{\varphi}, \boldsymbol{\upsilon}_{ki}\right) d\boldsymbol{\upsilon}_{ki}, \text{ with } w_{ki} = \left(N_k - 1\right) \frac{\overline{ms}_{ki}}{\sum_{i=1}^{N_k - 1} \overline{ms}_{ki}},$$

where $\overline{ms}_{ki}$ is the average market share of brand i in category k over the observation period. Since the likelihood function includes a multidimensional integral that does not provide a closed form solution it must be evaluated by using approximation methods. We follow the procedure suggested by Vermunt and Magidson (2005) and apply Gauss-Hermite numerical integration. Following Bock and Aitkin (1981), the multidimensional integral is replaced by multiple sums and approximated by:

$$\int_{\boldsymbol{\upsilon}_{ki}} f\left(\boldsymbol{\upsilon}_{ki}\right) f\left(\mathbf{y}_{ki} \middle| \mathbf{x}_{ki}, \boldsymbol{\Theta}, \boldsymbol{\varphi}, \boldsymbol{\upsilon}_{ki}\right) d\boldsymbol{\upsilon}_{ki} \approx \sum_{b_1=1}^{B} \sum_{b_2=1}^{B} f\left(\mathbf{y}_{ki} \middle| \mathbf{x}_{ki}, \boldsymbol{\Theta}, \boldsymbol{\varphi}, \upsilon_{b_1}, \upsilon_{b_2}\right) P_{b_1} P_{b2}, \tag{W.4}$$

where B denotes the number of quadrature nodes per dimension, υ_{b_1} and υ_{b_2} are the locations and P_{b_1} and P_{b_2} are the weights corresponding to quadrature nodes b_1 and b_2. To keep the estimation with two random variables manageable we set B = 5. The procedure makes use of nodes and weights as provided in quadrature tables by Stroud and Secrest (1966).

The maximization problem of the likelihood function is solved by employing the EM and the Newton-Raphson algorithms in combination. The advantage of EM is its stability in approaching the optimum whereas Newton-Raphson is faster than EM when it is close to the optimum. We start with a number of 250 EM iterations at maximum and switch to the Newton-Raphson algorithm to obtain the final solution. We use 500 random sets of start parameters to minimize the danger of finding a local optimum.

Our estimation approach enables us also to obtain brand-specific parameters:

$$\hat{\theta}_{lki} = \hat{\bar{\theta}}_l + \hat{\lambda}_{l1} \hat{E}\left(\upsilon_{1ki} \middle| \mathbf{x}_{ki}, \mathbf{y}_{ki}\right) + \hat{\lambda}_{l2} \hat{E}\left(\upsilon_{2ki} \middle| \mathbf{x}_{ki}, \mathbf{y}_{ki}\right), \tag{W.5}$$

where $\hat{\mathrm{E}}\left(\upsilon_{1ki} \middle| \mathbf{x}_{ki}, \mathbf{y}_{ki}\right)$ and $\hat{\mathrm{E}}\left(\upsilon_{2ki} \middle| \mathbf{x}_{ki}, \mathbf{y}_{ki}\right)$ are posterior means of the latent random variables that are computed using Gauss-Hermite quadrature. For example, we have

$$\begin{aligned}\hat{\mathrm{E}}\left(\upsilon_{1ki} \middle| \mathbf{x}_{ki}, \mathbf{y}_{ki}\right) &= \frac{\int_{-\infty}^{\infty} \upsilon_{1ki} \mathrm{f}\left(\mathbf{y}_{ki} \middle| \mathbf{x}_{ki}, \upsilon_{1ki}\right) \mathrm{d}\upsilon_{1ki}}{\int_{-\infty}^{\infty} \mathrm{f}\left(\mathbf{y}_{ki} \middle| \mathbf{x}_{ki}, \upsilon_{1ki}\right) \mathrm{d}\upsilon_{1ki}} \\ &\approx \frac{\sum_{b_1=1}^{B} \upsilon_{b_1} \mathrm{f}\left(\mathbf{y}_{ki} \middle| \mathbf{x}_{ki}, \upsilon_{b_1}\right) \mathrm{P}_{b_1}}{\sum_{b_1=1}^{B} \mathrm{f}\left(\mathbf{y}_{ki} \middle| \mathbf{x}_{ki}, \upsilon_{b_1}\right) \mathrm{P}_{b_1}}.\end{aligned} \tag{W.6}$$

ESTIMATION RESULTS

Table W1 displays the full estimation results of the brand sales model including the effects of covariates and distributions of random parameters. Estimates in this table are not corrected for temporal aggregation bias, yet. These results are reported in the paper (table 4).

Table W1 Parameter Estimates of Brand Sales Model (2,831 Brands)

	Parameter estimate		*Estimated std. dev. of parameter distribution*	
Constant	**4.32**	(.042)	**1.50**	(.024)
Lagged own brand sales	**.712**	(.004)	**.105**	(.003)
Own marketing mix				
Detailing	**.061**	(.001)	**.046**	(.001)
Professional journal advertising	**.027**	(.001)	**.019**	(.001)
DTC	**.039**	(.001)	**.022**	(.001)
Price	**-.215**	(.003)	**.090**	(.003)
Competitive marketing mix				
Detailing (1st entrant)	**.011**	(.001)	**.015**	(.001)
Detailing (2nd entrant)	**-.010**	(.001)	.002	(.001) NS
Detailing (late entrants)	-.002	(.002) NS	**.015**	(.001)
Professional journal advertising (1st entrant)	**.013**	(.002)	**.012**	(.002)
Professional journal advertising (2nd entrant)	**-.017**	(.002)	**.009**	(.002)
Professional journal advertising (late entrants)	**.006**	(.001)	.001	(.001) NS
DTC (1st entrant)	-.003	(.004) NS	.007	(.004) NS
DTC (2nd entrant)	**-.011**	(.003)	**.008**	(.004)
DTC (late entrants)	**.005**	(.001)	**.004**	(.001)
Price (innovative drugs)	**-.095**	(.006)	**.041**	(.003)
Price (generic/me-too drugs)	**.021**	(.004)	**.032**	(.002)
Innovativeness (reference group = generic/me-too drugs)				
Incremental innovation (0/1)	**.192**	(.009)		
Market breakthrough (0/1)	**.428**	(.017)		
Technological breakthrough (0/1)	**-.057**	(.009)		
Radical innovation (0/1)	**.586**	(.010)		
Covariates				
Elapsed time since launch: parameter γ_1	**-.006**	(.001)	**.009**	(.001)
Elapsed time since launch: parameter γ_2	**.088**	(.008)	**.053**	(.008)
Order of entry	**-.094**	(.007)		
New drug indication(s) (0/1)	**1.07**	(.040)	**1.61**	(.034)
Loss of patent protection (0/1)	**-.474**	(.015)	**.665**	(.016)
Introduction of a radical innovation (0/1)	**.041**	(.016)		
Log Likelihood = -20,956.80/ Sample size = 47,308				

Notes: Standard errors in parentheses; **Bold** = significant, NS = Not significant (p > .05; two-sided t-test); Parameter estimates for seasonal dummies and category dummies are not shown but available from the authors upon request.

Temporal Bias Correction

Tellis and Franses (2006) demonstrate that the unit exposure time, which they define as the time interval between two advertising exposures, is the optimal data interval. If data are not sampled at this interval a correction procedure is necessary to account for temporal aggregation bias in short-term effects. A bias does not exist if data are too disaggregated. Given that physicians are rarely visited by the same sales representative and for the same product more than four times a year and that price promotion activities are not typical for this industry, we do not expect biases in the detailing and price coefficients. However, we are more concerned about aggregation biases in the professional journal advertising and DTC coefficients and therefore apply the correction procedure as proposed by Tellis and Franses (2006).

Define M=3 months as the number of microperiods underlying quarterly aggregation and r=1 as the microtime when the marketing impulse occurs first in each aggregate (quarterly) period. Let τ denote the correction parameter, which is defined as $\left(1+\kappa+\cdots+\kappa^{M-r}\right)$, where κ is the M-root of κ^{M}, the carryover coefficient of the aggregated brand sales model. Hence, in our case, we have $\kappa^{M}=\hat{\delta}$ (the estimated carryover coefficient of table W1) and therefore $\hat{\tau}=1+\hat{\delta}^{1/3}+\hat{\delta}^{2/3}$. Let α measure the effect of a marketing mix element. We obtain the corrected short-term effect by (see Equation 16 in Tellis and Franses 2006, p. 221):

$$\hat{\alpha}_{\text{short-term}}^{\text{corr}}=\frac{\hat{\alpha}}{1+\hat{\delta}^{1/3}+\hat{\delta}^{2/3}}. \tag{W.7}$$

We approximate its standard error by application of the delta method:

$$\mathrm{SE}\left(\hat{\alpha}_{\text{short-term}}^{\text{corr}}\right)=\sqrt{\frac{1}{\tau^{2}}\mathrm{Var}\left(\hat{\alpha}\right)+\frac{\hat{\alpha}^{2}}{\tau^{4}}\left(\frac{1}{3\hat{\delta}^{2/3}}+\frac{2}{3\hat{\delta}^{1/3}}\right)^{2}\mathrm{Var}\left(\hat{\delta}\right)+\frac{2\hat{\alpha}}{\tau^{3}}\left(\frac{1}{3\hat{\delta}^{2/3}}+\frac{2}{3\hat{\delta}^{1/3}}\right)\mathrm{Cov}\left(\hat{\alpha},\hat{\delta}\right)}. \tag{W.8}$$

There is no need to correct long-term effect estimates. We obtain the long-term effect by:

$$\hat{\alpha}_{\text{long-term}}=\frac{\hat{\alpha}}{1-\hat{\delta}}. \tag{W.9}$$

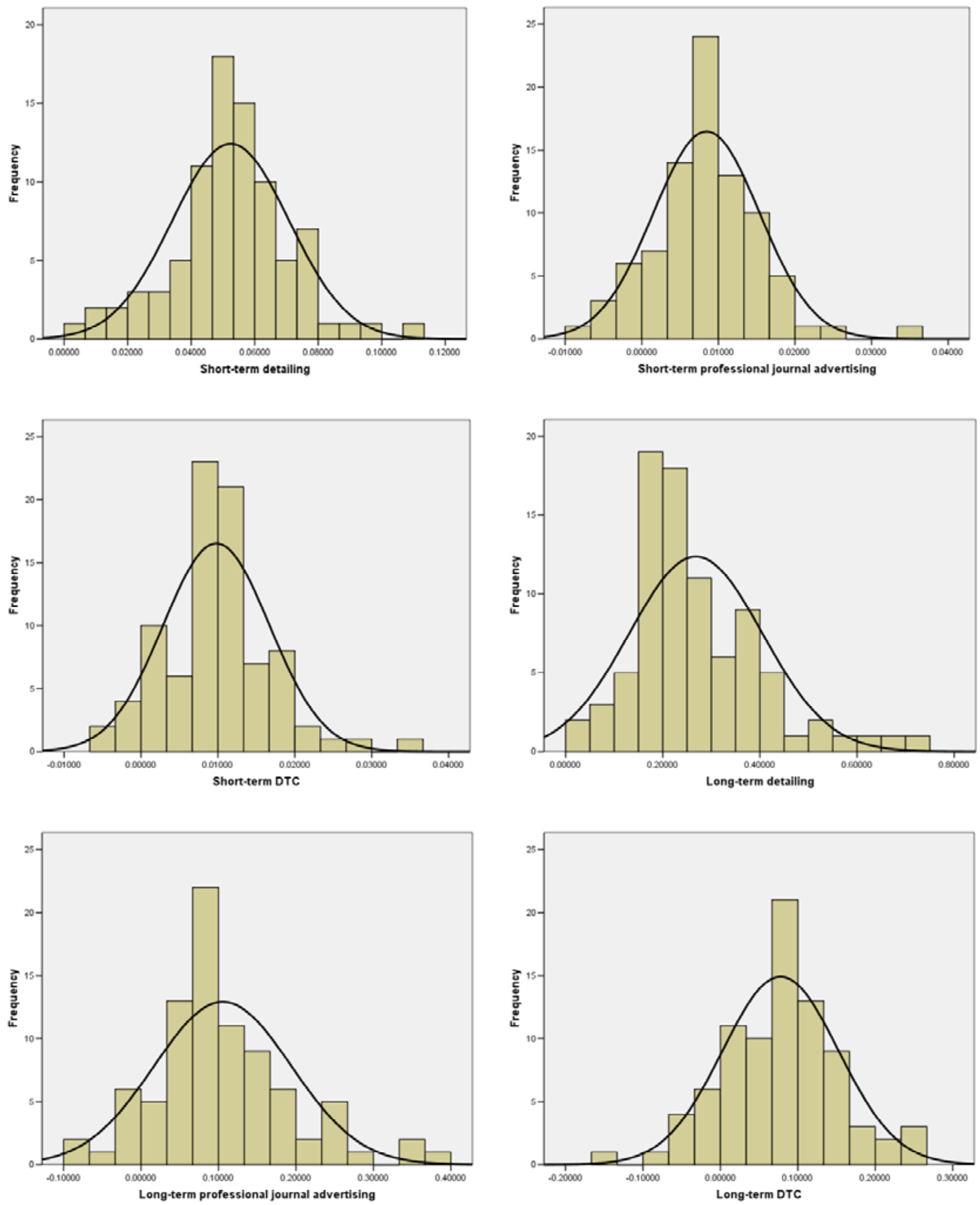
Standard errors are again approximated by using the delta method:

$$\mathrm{SE}\left(\hat{\alpha}_{\text{long-term}}\right)=\sqrt{\frac{1}{\left(1-\hat{\delta}\right)^{2}}\mathrm{Var}\left(\hat{\alpha}\right)+\frac{\hat{\alpha}^{2}}{\left(1-\hat{\delta}\right)^{4}}\mathrm{Var}\left(\hat{\delta}\right)+\frac{2\hat{\alpha}}{\left(1-\hat{\delta}\right)^{3}}\mathrm{Cov}\left(\hat{\alpha},\hat{\delta}\right)}. \tag{W.10}$$

Primary Demand Effects

Figure W1 displays the distribution of primary demand effects obtained by Equation (5).

Figure W1 Distribution of Competition-Neutral Primary Demand Elasticities



TESTING MODEL ASSUMPTIONS

Marketing Dynamics

The dynamics assumed in our brand sales model are consistent with the assumptions of the partial adjustment model. The well-known Koyck-model represents another alternative to model marketing dynamics. Both models lead to almost identical linear estimation equations that include lagged sales and other predictor variables. The only difference lies in the structure of the error term. The Koyck-model introduces serial correlation into the error term that is determined by the carryover coefficient. In the partial adjustment model, errors are assumed to be i.i.d. Following Johnston (1984, 347ff), we tested whether the restrictions of the Koyck-model hold for our data. Since we do not find serial correlation ($\rho = .068$, $p > .05$), we need to reject this assumption and conclude that the partial adjustment model is the more appropriate specification. We also tested a Koyck-type model that assumes heterogeneous carryover parameters (Johnston 1984; Mela, Gupta, and Lehmann 1997). This specification puts even more restrictions on model parameters and the error term (see Johnston 1984, 347). Specification tests revealed that our data are not consistent with these restrictions. Specifically, *F*-tests on linear parameter restrictions, which result from the Koyck-specification, were typically rejected at a high level ($p<.01$).

We also considered estimating a model that is free of parametric restrictions on the lag structure. Such a model involves additional lags. While it is attractive to specify an unrestricted lag structure and let the data speak, the practical application of this approach is often plagued by collinearity between the regressors, leading to great imprecision of the associated estimates (see also Johnston 1984, 352). When estimating a model that includes one additional lag we already find many estimates carrying high standard errors. More importantly, we compared this specification with our proposed specification and found that the Bayesian Information Criterion favors the proposed specification (BIC=24,548) over the alternative specification (BIC=24,622).

Heterogeneity

Our heterogeneity specification (see again expression W.2) allows for correlation of random parameters across different marketing mix elements. We compare this specification with a restricted specification of the variance-covariance matrix that assumes uncorrelated brand sales effects. Hence, the off-diagonal elements of (W.2) are set to zero. Our proposed specification appears to be a better representation of the data generating process ($BIC_{proposed} = 43{,}147$ vs. $BIC_{uncorrelated} = 52{,}760$).

Endogeneity

Marketing expenditures might be endogenous introducing a correlation between predictor variables and the error term that leads to biased or inconsistent estimates. We are not concerned about a potential *cross-sectional* correlation since we control for that by including a random brand constant. However, there may be concerns about an error correlation of expenditure variables over time. To investigate this issue we apply the Hausman-test to the model in first differences (Greene 2004). Anderson and Hsiao (1982) proof that values of the variables that are at least two periods lagged are uncorrelated with the error term by construction in a first-difference model and therefore serve as valid instruments. Using these instruments we do not find evidence that marketing decision variables are endogenous in our dataset [$\chi^2(4) = 4.83$, $p > .30$]. All first stage regressions have high R^2 and *F*-values far above 10 indicating that the instruments are not weak (Stock, Wright, and Yogo 2002).

REFERENCES

Anderson, T. W. and Cheng Hsiao (1982), "Formulation and Estimation of Dynamic Models Using Panel Data", *Journal of Econometrics*, 18, 47-82.

Bock, R. D. and M. Aitkin (1981), “Marginal Maximum Likelihood Estimation of Item Parameters,” *Psychometrika*, 46, 443-459.

Greene, William H. (2004), *Econometric Analysis*, 5th ed., Upper Saddle River: Prentice Hall.

Johnston, J. (1984), *Econometric Methods*. 3rd. ed., New York: McGraw Hill.

Mela, Carl F. Gupta, Sunil, and Donald R. Lehmann (1997), “The Long-Term Impact of Promotion and Advertising on Consumer Brand Choice,” *Journal of Marketing Research*, 34, 248-261.

Stock, J. H., J. H. Wright, and M. Yogo (2002), "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics*, 20, 518-529.

Stroud, A. H. and D. Secrest (1966), *Gaussian Quadrature Formulas*. Englewood Cliffs: Prentice Hall.

Tellis, Gerard J. and Philip Hans Franses (2006), “Optimal Data Interval for Estimating Advertising Response,” *Marketing Science*, 25, 217-229.

Vermunt, Jeroen K. and Kay Magidson (2005), *Technical Guide for Latent GOLD 4.0: Basic and Advanced*, Belmont Massachusetts: Statistical Innovations Inc.