Web Appendix

The Sense and Non-sense of Holdout Sample Validation in the Presence of Endogeneity

In part A of this Web Appendix we present a sample of papers that use holdout sample validation in the presence of endogeneity. In parts B-F of this Web Appendix we derive the results discussed in sections 2 and 3 of the main document. Parts G and H provide the results of two simulation studies that consider a situation with limited dependent variables and a situation with response heterogeneity.

A. Extended literature review

Table A.A1: Sample of studies that use holdout sample validation and endogeneity correction jointly

Authors	**Endogeneity issue**	**Relevant Text Extracted from Article**
Ailawadi et al. (2005)	Price Deal	"We compare the predictive ability of our model to two benchmarks." (p.18) "A reaction-function-based model that assumes competitors and the retailer will continue to react to each other's decisions the same way they have in the past, even in the wake of a major policy shift, has some predictive ability but is clearly dominated by our model. Also, a dynamic model that does not consider retailer response performs quite poorly." (p. 22)
Andrews and Currim (2009)	Price	"For both product categories studied, the model specified with random coefficient distributions for covariates and random effect distributions for common demand shocks, estimated using an instrumental variables approach to control for endogeneity, produced superior fit and predictive accuracy. In particular, the fully specified model dominated all finite mixture and random coefficients benchmark models, including those without common demand shocks and/or corrections for endogeneity." (p. 204)
Bass et al. (2007)	Price	"The evidence clearly supports the predictive performance of our proposed model." Model 1 Full model; Model 2 No advertising interactions; Model 3 Aggregate GRPs; Model 4 Square root GRP Model 5 Linear GRP (p. 187)
Besanko et al. (1998)	Price	"This mixed performance of the proposed model in predicting brand shares is not entirely unexpected, since the model simultaneously predicts prices as well, while the benchmark model (SUR estimator) takes prices as exogenously given. In addition, other studies have found that simultaneous equations models in which parameter estimates have been obtained with OLS often predict just as well as when parameters are estimated through more sophisticated methods, even though the OLS estimates are biased (Kennedy 1992)." (p. 1542)
Briesch et al. (2008)	Zero Brand Sales	"We showed that both model forms perform better than the MNL model. In general, the selection model (M1) is preferable on two counts. First, this model form provided the best in-sample and out-of-sample fits and hit rates." (p. 631)
Chan et al. (2007)	Location Price	"We also estimate a benchmark model (called BENCH) that restricts potential gasoline demand to be equal to the local population (e.g., Seim 2006). To understand how well we are able to predict observed gasoline station locations using our chosen set of six demographic variables, we compare the predictive ability of PROPOS with that of BENCH. The QMSE based on PROPOS is 7210, and that based on BENCH is 15,199, thus indicating more than 50% predictive gains from using demographic variables to predict local potential gasoline demand, as in our proposed location model." (p. 630)
Chintagunta (2001)	Price	"We also carried out a predictive validation exercise on four holdout weeks for the logit and probit outside good models. […] The mean absolute percentage error from the logit model is 27% and that for the probit model is 23%." (p. 454)
Chintagunta and Dubé (2005)	Price	"As a benchmark, we use the results from the model that uses the individual household data and the MLE method. This is the random coefficients choice model but without the unobserved attribute. The benchmark model yielded an MAPD measure of .51; for the proposed model the value was .33. Thus, the proposed model outperforms the

		benchmark model on the forecasting criterion." (p. 374)
Danaher et al. (2008)	Advertising	"We compare several models, ranging from a benchmark model with fixed own- and cross-brand advertising elasticities to models that accommodate various forms of competitive interference. […] Table 2 provides the MSE and MAD values for the benchmark model and the two models that incorporate competitive clutter, as given in Equations 7 and 8. It shows that the exponential and logistic models perform the best for both the calibration and validation data and for all categories." (p. 218)
Dong et al. (2009)	Detailing	"We also compared the two approaches using the holdout sample test. Specifically, we compute the likelihood value for the final quarter at each draw in the MCMC process for both approaches (Bodapati 2008). We then compute the posterior mean of the log-likelihood across all the simulation draws—this is -6184 for the approach that ignores strategic firm behavior and -5707 for the proposed approach. These results suggest that the proposed approach captures the data-generating process much better." (p. 219)
Drèze et al. (2004)	Price Promotion	"To test for possible over-fitting of the model, we ran an out-of-sample validation" […] "For instance, the MSE for the expenditure model was 0.32 in-sample and 0.36 out-of-sample." (p. 82)
Fischer and Albers (2010)	Lagged Sales	"We compare the performance of the proposed brand level model with that of the traditional category sales model." (p. 114) "The proposed brand-level model better predicts category sales changes in both the estimation and the holdout sample." (p. 116)
Kumar et al. (2009)	Distribution	"We also used the new data to validate our model, and a MAPE is calculated for each model." (p. 653)
Kumar et al. (2008)	Time-varying missing variables	"The Appendix provides the in-sample fit capability and the predictive capability of the proposed joint model of category choice and purchase timing along with other benchmark models. The in-sample fit (Table A1) and predictive accuracy (Table A2) results provide support for the full model specification outlined in Equation 5." (p. 60)
Leenheer et al. (2007)	Loyalty Program Membership	"Overall, we find that the model with endogeneity correction predicts the changes in share-of-wallet due to membership changes considerably better than the naïve benchmark model that does not accommodate endogeneity." (p. 42)
Li et al. (2009)	Auction entry	"In addition, the proposed model performs better than the two benchmark models on all other dimensions for both the estimation sample and the holdout sample, which indicates the importance of recognizing latent bidder competition with endogenous entry and bidder preference heterogeneity." (p. 86)
Manchanda et al. (2004)	Detailing	"In this section, we compare the conditional NBD hierarchical model with the full hierarchical model in both time-series and cross-sectional validation experiments. Both the time-series and the cross-sectional validation exercises provide evidence that our full modeling approach improves prediction." (p. 475)
Narayanan et al. (2004)	Price; Detailing; DTC	"Second, we also compared predictions for a holdout sample with the sets of instruments. The overall results are consistent;" (p. 96)
Neslin (1990)	Coupon Redemption	"The model was compared to a naïve model that simply regressed market share versus the trend and brand dummy variables. […] Table 6 shows that, on average, the techniques have equal forecasting ability." (p. 139)
Richards (2007)	Price Share of Products on Sales	"Finally, we calculate Theil's *U* over the final twenty weeks of data in order to evaluate the ability of the model to forecast accurately over a validation data set, both for the individual equations and the model as a whole. As the results in Table 3 show, the *U* statistic for the share equation suggests a high degree of predictive accuracy while the measures for the margin and proportion of sale products are slightly higher." (p. 79)

Song and Chintagunta (2006)	Price	"Importantly, we also account for potential correlation between the prices and the error term in the log-log model by instrumenting for prices." (p. 1605) "Hence, it appears that on all criteria considered, the proposed model predicts better than the log-log model." (p. 1607)
Van Dijk et al. (2004)	Shelf Space	*"SPATIAL-CHAR* [is] the proposed model that corrects for endogeneity with a spatial structure based on similarities between store profiles." (p. 262) "We note that for *each* of the brands, the *SPATIAL-CHAR* model outperforms *OLS* in predictive validity." (p. 275)
Van Nierop et al. (2008)	Facings	"We compare our model against a per-item regression and achieve better results for all diagnostics, in particular out of sample." (p. 1077)
Venkatesan and Kumar (2004)	Antecendents of Purchase Frequency and Contribution Margin	"The RAE is given by the ratio of the model MAD to the MAD based on the moving average measure. Based on the RAE measure, the generalized gamma model with time-varying covariates (Model 3) has an RAE of .51, compared with that of a naïve moving average technique." (p. 123)
Venkatesan et al. (2007)	Antecedents of Purchase Quantity	"The RAE measures indicate that all the models (the base, the proposed, and the benchmark models) provide better predictive accuracy than simple heuristics (given that the RAEs are less than 1), in addition to providing a framework for linking customer behavior to marketing actions." (p. 588)
Zhu et al. (2009)	Market Structure	"For the holdout sample, the overall hit rate was 67%, which is reasonable given that our predicted market structure is based on identities of firms and not just the total number." (p. 463)

B. Asymptotic Mean Squared Prediction Errors for OLS versus IV using X only

The ordinary least squares (OLS) estimator for β in model (1) in the main document based on the first N_1 observations is given by

$$\hat{\beta}_{OLS} = (X_1'X_1)^{-1}X_1'Y_1 = \beta + (X_1'X_1)^{-1}X_1'\epsilon_1. \tag{B1}$$

Under the assumption that the regressors and errors are independent, such that $E(\epsilon_1|X_1) = 0$, the OLS estimator $\hat{\beta}_{OLS}$ is unbiased and consistent for β (e.g., Greene 2003, Chapter 5). When the regressors are endogenous ($E(\epsilon_1|X_1) \neq 0$), the instrumental variables (IV) approach may be used to obtain a consistent estimator for β. Let Z_1 be the N_1xL matrix containing the IVs, where $L \geq K$ for identification. The IV estimator for β is given by (e.g., Greene 2003, Chapter 5):

$$\hat{\beta}_{IV} = \left(X_1'P_{Z_1}X_1\right)^{-1}X_1'P_{Z_1}Y_1 = \beta + \left(X_1'P_{Z_1}X_1\right)^{-1}X_1'P_{Z_1}\epsilon_1, \tag{B2}$$

where $P_{Z_1} = Z_1(Z_1'Z_1)^{-1}Z_1'$. Under the usual assumption that the instruments are valid (e.g., Davidson and MacKinnon 1993), it can be shown that $\hat{\beta}_{IV}$ is a consistent estimator for β.

Using the OLS fitted model from equation (1) in the main document and the holdout predictors X_2, the predicted values are obtained as $\hat{Y}_{OLS,2} = X_2\hat{\beta}_{OLS}$. The holdout prediction errors are given by

$$\begin{aligned} \hat{e}_{OLS,2} &= Y_2 - \hat{Y}_{OLS,2} = X_2\beta + \epsilon_2 - X_2\beta - X_2(X_1'X_1)^{-1}X_1'\epsilon_1 \\ &= \epsilon_2 - X_2(X_1'X_1)^{-1}X_1'\epsilon_1, \end{aligned} \tag{B3}$$

and the sum of the squared errors is

$$\begin{aligned} \hat{e}_{OLS,2}'\hat{e}_{OLS,2} = {} & \epsilon_2'\epsilon_2 - \epsilon_2'X_2(X_1'X_1)^{-1}X_1'\epsilon_1 - \epsilon_1'X_1(X_1'X_1)^{-1}X_2'\epsilon_2 \\ & + \epsilon_1'X_1(X_1'X_1)^{-1}X_2'X_2(X_1'X_1)^{-1}X_1'\epsilon_1. \end{aligned} \tag{B4}$$

We now investigate the large sample properties of the sum of the squared errors in (B4), which may be seen as the "asymptotic" mean-squared prediction error, $\text{A.MSE}_{\text{OLS}}$ for short, for which we let both the estimation and holdout samples grow, i.e.,[1]

$$\text{A. MSE}_{\text{OLS}} = \text{ plim } \hat{e}'_{OLS,2}\hat{e}_{OLS,2} \;/\; N_2 \tag{B5}$$

$$= \sigma_\epsilon^2 - \Sigma_{\epsilon\psi,2}\text{plim}(X_1'X_1/N_1)^{-1}\Sigma_{\psi\epsilon,1} - \Sigma_{\epsilon\psi,1}\text{plim}(X_1'X_1/N_1)^{-1}\Sigma_{\psi\epsilon,2}$$

$$+ \Sigma_{\epsilon\psi,1}\text{plim}(X_1'X_1/N_1)^{-1}\text{plim}(X_2'X_2/N_2)\text{plim}(X_1'X_1/N_1)^{-1}\Sigma_{\psi\epsilon,1}.$$

It follows from Khintchine's theorem (e.g., Greene 1993, p. 365) with $E(X_i'\epsilon_i) = E(\Psi_i'\epsilon_i) = \Sigma_{\psi\epsilon}$, that plim $X'\epsilon/N = \Sigma_{\psi\epsilon}$. To simplify (B5), we assume for the moment that the sample is split randomly into an estimation and holdout sample and hence $\Sigma_{\epsilon\psi,1} = \Sigma_{\epsilon\psi,2} = \Sigma_{\epsilon\psi}$,[2] and that the regressors X_1 and X_2 have the same moments, so that $\text{plim}(X_1'X_1/N_1) = \text{plim}(X_2'X_2/N_2) = Q_{XX}$. Now (B5) simplifies to

$$\text{A.MSE}_{\text{OLS}} = \sigma_\epsilon^2 - \Sigma_{\epsilon\psi}Q_{XX}^{-1}\Sigma_{\psi\epsilon}. \tag{B6}$$

Hence, the asymptotic mean squared prediction error for OLS is smaller than the variance of the errors, σ_ϵ^2, because $\Sigma_{\epsilon\psi}Q_{XX}^{-1}\Sigma_{\psi\epsilon}$ is nonnegative.

When instead of the OLS fitted model, the IV fitted model is used to predict the dependent variable in the holdout sample, $\hat{Y}_{IV,2} = X_2\hat{\beta}_{IV}$, the prediction errors are given by

$$\hat{e}_{IV,2} = Y_2 - \hat{Y}_{IV,2} = X_2\beta + \epsilon_2 - X_2\beta - X_2\left(X_1'P_{Z_1}X_1\right)^{-1}X_1'P_{Z_1}\epsilon_1 \tag{B7}$$

$$= \epsilon_2 - X_2\left(X_1'P_{Z_1}X_1\right)^{-1}X_1'P_{Z_1}\epsilon_1,$$

and the squared prediction errors are

[1] We assume that the regressors X have finite moments. The result in (B5) follows from Slutsky's theorems (e.g., Ferguson 1996).

[2] We will relax the assumption $\Sigma_{\epsilon\psi,1} = \Sigma_{\epsilon\psi,2}$ below, which may be the case when the estimation and holdout samples are not split randomly.

$$\hat{e}'_{IV,2}\hat{e}_{IV,2} = \epsilon'_2\epsilon_2 - \epsilon'_2 X_2\left(X'_1 P_{Z_1} X_1\right)^{-1} X'_1 P_{Z_1}\epsilon_1 - \epsilon'_1 P_{Z_1} X_1\left(X'_1 P_{Z_1} X_1\right)^{-1} X'_2\epsilon_2 + \epsilon'_1 P_{Z_1} X_1\left(X'_1 P_{Z_1} X_1\right)^{-1} X'_2 X_2\left(X'_1 P_{Z_1} X_1\right)^{-1} X'_1 P_{Z_1}\epsilon_1. \quad \text{(B8)}$$

Asymptotically, the mean squared prediction errors are equal to[3]

$$\begin{aligned} \text{A.MSE}_{\text{IV}} &= \text{plim}\ \hat{e}'_{IV,2}\hat{e}_{IV,2} / N_2 \\ &= \sigma^2_\epsilon - \Sigma_{\epsilon\psi,2}\text{plim}\left(X'_1 P_{Z_1} X_1/N_1\right)^{-1}\text{plim}\ (X'_1 Z_1/N_1)\ \text{plim}\ (Z'_1 Z_1/N_1)^{-1}\mathbf{0} \\ &\quad - \mathbf{0}'\text{plim}\ (Z'_1 Z_1/N_1)^{-1}\text{plim}\ (Z'_1 X_1/N_1)\text{lim}\left(X'_1 P_{Z_1} X_1/N_1\right)^{-1}\Sigma_{\psi\epsilon,2} \\ &\quad + \mathbf{0}'\text{plim}\ (Z'_1 Z_1/N_1)^{-1}\text{plim}(\dots)\text{plim}\ (Z'_1 Z_1/N_1)^{-1}\mathbf{0} \\ &= \sigma^2_\epsilon, \end{aligned} \quad \text{(B9)}$$

where plim $Z'_1\epsilon_1/N_1 = 0$ because the instruments are exogenous (we suppressed the product matrices in the before-last term for notational simplicity). Thus, IV obtains an asymptotic MSE of σ^2_ϵ.

Now consider different extents of endogeneity in the estimation and holdout samples, i.e., $\Sigma_{\epsilon\psi,1} \neq \Sigma_{\epsilon\psi,2}$. If the regressors X_1 and X_2 have the same moments, so that $\text{plim}(X'_1 X_1/N_1) = \text{plim}(X'_2 X_2/N_2) = Q_{XX}$, (B9) remains the same, and (B5) simplifies to:

$$\text{A.MSE}_{\text{OLS}} = \sigma^2_\epsilon - \Sigma_{\epsilon\psi,2} Q^{-1}_{XX}\Sigma_{\psi\epsilon,1} - \Sigma_{\epsilon\psi,1} Q^{-1}_{XX}\Sigma_{\psi\epsilon,2} + \Sigma_{\epsilon\psi,1} Q^{-1}_{XX}\Sigma_{\psi\epsilon,1}. \quad \text{(B10)}$$

Similar expressions to (B6) and (B9) may be obtained for *in-sample* predictions using Y_1 and X_1. Consequently, standard model selection criteria based on fit measures, such as the R^2, tend to favor the OLS model over the IV model (e.g., Verbeek 2008, p. 150).

[3] We assume that the usual (asymptotic) conditions for IV estimation are satisfied (e.g., Davidson and MacKinnon 1993), i.e., that the instruments have finite moments and have some correlation with *X* in the limit. Also, we assume that the instruments are exogenous.

C. Using $\hat{X}$ to predict in a holdout sample

We now analyze the fitted model that uses the predicted values for *X* to predict *Y* in the holdout sample.

An equivalent representation of the structural form of the IV model in (1) and (2) in the main document is

$$Y = (Z\Gamma)\beta + u \tag{C1}$$

$$X = Z\Gamma + \Upsilon$$

where the variance of the joint errors $u_i = \Upsilon_i\beta + \epsilon_i$ and Υ_i is given by

$$\Omega = \begin{bmatrix} \omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} = \begin{bmatrix} 1 & \beta' \\ 0 & I \end{bmatrix} \Sigma \begin{bmatrix} 1 & 0 \\ \beta & I \end{bmatrix}. \tag{C2}$$

(C1) may be used to analyze the fitted model that uses the predicted values for *X* to forecast *Y* in the holdout sample, i.e.,

$$\hat{Y}_{\mathrm{IV}.\hat{X},2} = \hat{X}_2\hat{\beta}_{\mathrm{IV}}, \tag{C3}$$

where $\hat{X}_2 = Z_2\widehat{\Gamma}_{\mathrm{IV}}$. Using properties of the reduced form model in (C1), it follows that the asymptotic prediction errors converge to ω_{11} in (C2), i.e.,

$$\mathrm{A.MSE}_{\mathrm{IV}.\hat{X}} = \sigma_\epsilon^2 + \beta'\Sigma_{\nu\epsilon} + \Sigma_{\epsilon\nu}\beta + \beta'\Sigma_{\nu\nu}\beta. \tag{C4}$$

Whether or not the reduced form fitted model has lower mean-squared holdout prediction errors than the IV fitted model, depends on the term $d = \beta'\Sigma_{\psi\epsilon} + \Sigma_{\epsilon\psi}\beta + \beta'\Sigma_{\psi\psi}\beta$. As

$$d = \beta'\Sigma_{\psi\epsilon} + \Sigma_{\epsilon\psi}\beta + \beta'\Sigma_{\psi\psi}\beta = \beta'(2\Sigma_{\psi\epsilon} + \Sigma_{\psi\psi}\beta) \tag{C5}$$

it follows that $d = 0$ for $\beta = 0$ or $2\Sigma_{\psi\epsilon} + \Sigma_{\psi\psi}\beta = 0 \Leftrightarrow \beta = -2\Sigma_{\psi\psi}^{-1}\Sigma_{\psi\epsilon}$, so that $\mathrm{A.MSE}_{\mathrm{IV}.\hat{X}} = \mathrm{A.MSE}_{\mathrm{IV}}$. Since (C5) is a second-degree polynomial with a nonnegative coefficient for the square term of β, $\text{A.MSE}_{\text{IV}.\hat{X}} \leq \text{A.MSE}_{\text{IV}}$ for $\beta \in [-2\,\Sigma_{\psi\psi}^{-1}\,\Sigma_{\psi\epsilon}, 0]$, and $\text{A.MSE}_{\text{IV}.\hat{X}} >$ A.MSE_{IV} for $\beta \notin [-2\,\Sigma_{\psi\psi}^{-1}\,\Sigma_{\psi\epsilon}, 0]$.

Hence, the IV fitted model $\hat{Y}_{\text{IV}.\hat{X},2} = \hat{X}_2\hat{\beta}_{\text{IV}}$ may or may not give mean-squared errors that are lower than the mean-squared errors of the IV fitted model $\hat{Y}_{\text{IV},2} = X_2\hat{\beta}_{\text{IV}}$, depending on the specific values of the regression coefficients β and the elements in Σ.

D. Asymptotic Mean Squared Prediction Errors for OLS versus IV using *both* X *and* Z

This section studies whether holdout sample predictions may be improved by using both *X* and *Z*. We assume that holdout sample values for the instruments *Z* are available. Without making distributional assumptions, we can estimate regression model (1) [in the main text] consistently with IV. In the case of linear models, the IV estimates for β are equivalent to the estimates for β obtained from a control function approach (e.g., Petrin and Train 2010; Verbeek 2008, p. 144). The control function approach proceeds as follows:

(1) Regress X on Z and obtain the fitted residuals $\hat{\Psi} = X - Z\hat{\Gamma}$;

(2) Regress Y on X and $\hat{\Psi} = X - Z\hat{\Gamma}$.

The conditional mean and variance of *Y* given *X* and *Z* using the control function specification[4] is

$$E(Y|X,Z) = X\beta + (X - Z\Gamma)\phi = X\beta + (X - Z\Gamma)\Sigma_{\psi\psi}^{-1}\Sigma_{\psi\epsilon} \quad \text{(D1)}$$

$$\text{var}(Y|X,Z) = \sigma_\epsilon^2 - \phi'\Sigma_{\psi\psi}\phi = \sigma_\epsilon^2 - \Sigma_{\epsilon\psi}\Sigma_{\psi\psi}^{-1}\Sigma_{\psi\epsilon}. \quad \text{(D2)}$$

The holdout sample predictors (X_2, Z_2) can now be used to predict Y_2 using the fitted conditional mean:

[4] The control function approach splits up ϵ in (1) [main text] in a part that is correlated with *X* and a part that is not through the linear projection $\epsilon = \Psi\phi + \text{u}$, where Ψ correlates with *X*, e.g., see (2) [main document]. Here ϕ is a $K \times 1$ vector, u a $N \times 1$ vector with mean 0, and Ψ and u are uncorrelated by construction. It follows that $\text{var}(Y|X,Z) = \sigma_u^2$. From the expressions for $\sigma_\epsilon^2 = E\epsilon_i\epsilon_i$ and $\Sigma_{\epsilon\psi} = E(\epsilon_i\Psi_i)$ it follows that $\phi = \Sigma_{\psi\psi}^{-1}\Sigma_{\psi\epsilon}$ and $\sigma_\epsilon^2 = \phi'\Sigma_{\psi\psi}\phi + \sigma_u^2$, with $\sigma_u^2 = Eu_i^2$. Hence, $\text{var}(Y|X,Z) = \sigma_u^2 = \sigma_\epsilon^2 - \phi'\Sigma_{\psi\psi}\phi$.

$$\hat{Y}_{\text{IV.XZ},2} = X_2\hat{\beta}_{\text{IV}} + \left(X_2 - Z_2\hat{\Gamma}_{\text{IV}}\right)\hat{\Sigma}_{\psi\psi,\text{IV}}^{-1}\hat{\Sigma}_{\psi\epsilon,\text{IV}}, \tag{D3}$$

where the subscript 'XZ' indicates that Y_2 is predicted using holdout sample values of both *X* and *Z*. If the calibration and holdout samples are randomly split, then the unexplained variance in the holdout sample is simply $\text{var}(Y|X,Z)$, and the asymptotic mean squared prediction error is

$$\text{A.MSE}_{\text{IV.XZ}} = \sigma_\epsilon^2 - \Sigma_{\epsilon\psi}\Sigma_{\psi\psi}^{-1}\Sigma_{\psi\epsilon}. \tag{D4}$$

We now derive the asymptotic mean squared error for an OLS regression of *Y* on *X* and *Z*. Given *X* and *Z*, the dependent variable *Y* has mean (D1) and variance (D2), because an equivalent representation is an augmented regression model

$$Y = X\beta + (X - Z\Gamma)\Sigma_{\psi\psi}^{-1}\Sigma_{\psi\epsilon} + \text{U} = \text{X}\left(\beta + \Sigma_{\psi\psi}^{-1}\Sigma_{\psi\epsilon}\right) + \text{Z}\left(-\Gamma\Sigma_{\psi\psi}^{-1}\Sigma_{\psi\epsilon}\right) + \text{U} = \tag{D5}$$

$$= \text{X}\tilde{\beta}_1 + \text{Z}\tilde{\beta}_2 + \text{U},$$

where each element of U is independent with mean 0 and variance $\sigma_\epsilon^2 - \Sigma_{\epsilon\psi}\Sigma_{\psi\psi}^{-1}\Sigma_{\psi\epsilon}$, and *X* and *Z* are independent of U by construction. Hence, a standard OLS regression of *Y* on *X* and *Z* will estimate the (unrestricted) regression coefficients $\tilde{\beta} = (\tilde{\beta}_1', \tilde{\beta}_2')'$. If the instruments Z are valid, then $\tilde{\beta}_2 \neq 0$, unless *X* is exogenous. However, note that this "augmented" OLS approach does *not* correctly estimate the effect of *X* on Y, because $\tilde{\beta}_1 \neq \beta$, unless *X* is exogenous. Using the same reasoning as in deriving (D4), it follows that the "augmented" fitted model

$$\hat{Y}_{\text{OLS.XZ},2} = X_2\hat{\tilde{\beta}}_1 + Z_2\hat{\tilde{\beta}}_2 \tag{D6}$$

has the same asymptotic mean squared error in a holdout sample as (D4), i.e.,

$$\text{A.MSE}_{\text{OLS.XZ}} = \text{A.MSE}_{\text{IV.XZ}}. \tag{D7}$$

While this result hold without making any distributional assumptions, identical results can be obtained when the IV model is represented as a limited information simultaneous equation model with normally distributed error terms (Kleibergen and Zivot, 2003) and estimated with limited-information maximum likelihood (LIML), using standard results on multivariate normal distributions (e.g., Greene 2003).

Comparing $\text{A.MSE}_{\text{OLS}}$, A.MSE_{IV}, $\text{A.MSE}_{\text{IV.}\hat{X}}$, $\text{A.MSE}_{\text{OLS.XZ}}$, and $\text{A.MSE}_{\text{IV.XZ}}$

From (B6) we have $\text{A.MSE}_{\text{OLS}} = \sigma_\epsilon^2 - \Sigma_{\epsilon\psi} Q_{XX}^{-1} \Sigma_{\psi\epsilon}$. Furthermore, from (2) [main text],

$$X_i' X_i = (Z_i \Gamma + \Psi_i)'(Z_i \Gamma + \Psi_i) = \Gamma' Z_i' Z_i \Gamma + \Gamma' Z_i' \Psi_i + \Psi_i Z_i \Gamma + \Psi_i' \Psi_i. \quad \text{(D8)}$$

With the usual assumptions for the instruments (i.e., Γ is non-zero, the instruments have finite moments, and are exogenous),

$$E(X_i' X_i) = \Gamma' E(Z_i' Z_i) \Gamma + E(\Psi_i' \Psi_i), \quad \text{(D9)}$$

so that (Khintchine's theorem)[5]

$$Q_{XX} = \text{plim}(X'X/N) = \Gamma' Q_{ZZ} \Gamma + \Sigma_{\psi\psi}. \quad \text{(D10)}$$

Comparing $\text{A.MSE}_{\text{IV.XZ}}$ to $\text{A.MSE}_{\text{OLS}}$,

$$\text{A.MSE}_{\text{OLS}} - \text{A.MSE}_{\text{IV.XZ}} = \sigma_\epsilon^2 - \Sigma_{\epsilon\psi} Q_{XX}^{-1} \Sigma_{\psi\epsilon} - \left(\sigma_\epsilon^2 - \Sigma_{\epsilon\psi} \Sigma_{\psi\psi}^{-1} \Sigma_{\psi\epsilon}\right) = \quad \text{(D11)}$$

$$= \Sigma_{\epsilon\psi} \Sigma_{\psi\psi}^{-1} \Sigma_{\psi\epsilon} - \Sigma_{\epsilon\psi} Q_{XX}^{-1} \Sigma_{\psi\epsilon} = \Sigma_{\epsilon\psi} \left(\Sigma_{\psi\psi}^{-1} - Q_{XX}^{-1}\right) \Sigma_{\psi\epsilon}.$$

The matrix $\Gamma' Q_{ZZ} \Gamma$ in (D10) is positive semidefinite[6] so that[7] $Q_{XX} > \Sigma_{\psi\psi}$, which implies $\Sigma_{\psi\psi}^{-1} > Q_{XX}^{-1}$, or equivalently, that $\Sigma_{\psi\psi}^{-1} - Q_{XX}^{-1}$ is positive definite[8]. Hence[9], $\text{A.MSE}_{\text{OLS}} - \text{A.MSE}_{\text{IV.XZ}} > 0$ for all nonzero $\Sigma_{\epsilon\psi}$, and $\text{A.MSE}_{\text{IV.XZ}} \leq \text{A.MSE}_{\text{OLS}}$ (where equality holds for $\Sigma_{\epsilon\psi} = 0$ or $\Gamma = 0$). In sum,

$$\text{A.MSE}_{\text{IV.XZ}} = \text{A.MSE}_{\text{OLS.XZ}} \leq \text{A.MSE}_{\text{OLS}} \leq \text{A.MSE}_{\text{IV}}. \quad \text{(D12)}$$

[5] Note that Q_{XX} and Q_{ZZ} are not the (asymptotic) variance-covariance matrices of *X* and *Z*, but the second moments. These will equal the variance-covariance matrices if the first moments (mean) are zero.
[6] If *A* is positive definite, then any matrix of the form *B'AB* is positive definite if *B* has full column rank, and positive semidefinite otherwise (Davidson and MacKinnon 1993, p. 788).
[7] If *A* is positive definite and *B* is positive semidefinite, then *A*+*B*>*A* (Greene 2003, p. 49).
[8] Greene (2003, p. 49) and Davidson and MacKinnon (1993, p. 789).
[9] Apply the definition of positive definite matrices. A nxn matrix *A* is said to be positive definite if the quadratic form *x'Ax* is positive for all nonzero *n*-vector *x*, e.g. Davidson and MacKinnon (1993) p. 787.

Comparing $\text{A.MSE}_{\text{IV}.\hat{X}}$, note that in (C5) the first-order derivative of d with respect to β is,

$$\frac{\partial d}{\partial \beta} = 2\Sigma_{\nu\epsilon} + 2\Sigma_{\nu\nu}\beta, \quad \text{(D13)}$$

which is zero for $\beta = -\Sigma_{\nu\nu}^{-1}\Sigma_{\nu\epsilon}$. Substituting this result in (C4) gives $\text{A.MSE}_{\text{IV}.\hat{X}} = \sigma_\epsilon^2 - \Sigma_{\epsilon\nu}\Sigma_{\nu\nu}^{-1}\Sigma_{\nu\epsilon}$, so that (D12) becomes $\text{A.MSE}_{\text{IV}.\hat{X}} = \text{A.MSE}_{\text{OLS.XZ}} = \text{A.MSE}_{\text{IV.XZ}} \leq \text{A.MSE}_{\text{OLS}} \leq \text{A.MSE}_{\text{IV}}$ when $\beta = -\Sigma_{\nu\nu}^{-1}\Sigma_{\nu\epsilon}$. Hence, the IV fitted model $\hat{Y}_{\text{IV}.\hat{X},2} = \hat{X}_2\hat{\beta}_{\text{IV}}$ may or may outperform OLS or IV in a holdout sample validation task depending on β and Σ.

E. Holdout Sample Predictions for Exogenous Regressors

If the endogenous nature in the holdout sample is not preserved, for instance when the estimation and validation samples are not randomly split, then $\Sigma_{\epsilon\psi,1} \neq \Sigma_{\epsilon\psi,2}$. In the special case that X_2 is exogenous, $\Sigma_{\psi\epsilon,2} = \text{plim}\, X_2'\epsilon_2/N_2 = 0$. Although the A.MSE_{IV} in (B9) does not change, (B6) becomes

$$\text{A.MSE}_{\text{OLS}} = \sigma^2 + \Sigma_{\epsilon\psi,1}Q_{XX}^{-1}\Sigma_{\psi\epsilon,1}, \quad \text{(E1)}$$

and $\text{A.MSE}_{\text{IV}} \leq \text{A.MSE}_{\text{OLS}}$.

F. Pseudo-exogenous regressors for holdout predictions

We show that when data are observed in at least two dimensions and the endogeneity is concentrated in one dimension, a pseudo-exogenous holdout sample can be constructed to validate IV and OLS in a holdout sample prediction. Consider a situation where T observations are available for N subjects in a regression model, i.e.,

$$y_{it} = x_{it}\beta + u_{it}, \quad \text{(F1)}$$

where x_{it} is $1xK$ vector of regressors. Suppose that x_{it} is endogenous but endogeneity is in one dimension, e.g., $u_{it} = \alpha_i + \epsilon_{it}$ where x_{it} is correlated with α_i but not with $\epsilon_{it'}$ for all t'. Now,

$\text{Var}(u_i) = \Sigma$, with diagonal elements $\sigma_\alpha^2 + \sigma_\epsilon^2$ and off-diagonal elements σ_α^2. For the total sample, let $u = L\alpha + \epsilon$ where $L = I_N \otimes \iota_T$, and $\text{Var}(u) = I_N \otimes \Sigma = \sigma_\epsilon^2 \Theta$, where $\Theta = I_{NT} + (\sigma_\alpha^2 / \sigma_\epsilon^2) LL'$, and

$$Y = X\beta + L\alpha + \epsilon = X\beta + u, \quad \text{(F2)}$$

where Y is $NTx1$, X is $NTxK$, α is $Nx1$, and ϵ and u are $NTx1$ (i.e., the N subjects $i = 1, \ldots, N$ are stacked). Assume that the total sample $N_t = NT$ is split up in an estimation and holdout sample. Let the estimation sample be $i = 1, \ldots, N_1$ and $t = 1, \ldots, T_1$. Now three holdout samples may be obtained, as shown in Table A.E1.

Table A.E1
Different forms of holdout samples using multilevel data

Data matrix	$t = 1, \ldots, T_1$	$t = T_1 + 1, \ldots, T_1 + T_2$
$i = 1, \ldots, N_1$	Estimation sample (sample size $N_e = N_1 T_1$)	Holdout sample A (sample size $N_h = N_1 T_2$)
$i = N_1 + 1, \ldots, N_1 + N_2$	Holdout sample B (sample size $N_h = N_2 T_1$)	Holdout sample C (sample size $N_h = N_2 T_2$)

The total sample size is $N_t = N_e + N_h$, where N_e and N_h are given in the table. As before, Y_1 is an $N_e x1$ vector and Y_2 is an $N_h \times 1$ vector containing the dependent variable in the estimation and holdout samples, respectively, and are obtained by collecting the corresponding rows from Y (same for X and ϵ).

In (E2), X is endogenous because X correlates with u, and the results discussed in section 2.2 in the main document apply, regardless of whether the holdout sample is of type A, B, or C. The large sample properties for $\hat{\beta}_{\text{OLS}}$ and $\hat{\beta}_{\text{IV}}$ hold under fairly general conditions (e.g., Greene 2003, sec. 11.2) even if $\text{Var}(u)$ is nondiagonal. For the purpose of this study we need only be concerned about the point estimates of β (i.e., whether or not consistency holds) that are used for holdout sample predictions. In the following we are not concerned with the efficiency of the estimates and use $\hat{\beta}_{OLS}$ to indicate an estimator that is inconsistent (e.g., OLS or GLS), and use $\hat{\beta}_{IV}$ to indicate an estimator that is consistent (e.g., two- or three-stage least squares, or fixed effects estimation).

An opportunity to create a pseudo-exogenous holdout sample arises in this case by exploiting that endogeneity is in dimension i only. Here a transformation that produces observations in deviations from individual means eliminates the unobserved individual effects α_i. Let $Q = I_{NT} - L(L'L)^{-1}L'$ be the $NT \times NT$ matrix that produces deviations from individual means, then

$$QY = QX\beta + Qu = QX\beta + Q\epsilon \quad \text{(F3)}$$

or

$$\tilde{Y} = \tilde{X}\beta + \tilde{\epsilon} \quad \text{(F4)}$$

(i.e., the it-th row is equal to $y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)\beta + (\epsilon_{it} - \bar{\epsilon}_i)$, where $\bar{y}_i = 1/T \sum_t y_{it}$ etc.) and $\text{Var}(Q\epsilon) = Q\text{Var}(\epsilon)Q' = \sigma_\epsilon^2 Q$ (i.e., the diagonal elements are equal to $\sigma_\epsilon^2 - \sigma_\epsilon^2/T = \sigma_\epsilon^2\,(T-1)/T$). The transformed regressors QX are not correlated with the error term $Q\epsilon$. Predictions may now be made with the fitted model that only uses information in the 'pseudo-regressors' QX:

$$\hat{Y}_{\text{IV-pseudo,2}} = \tilde{X}_2\hat{\beta}_{\text{IV}}, \quad \text{(F5)}$$

and

$$\hat{Y}_{\text{OLS-pseudo,2}} = \tilde{X}_2\hat{\beta}_{\text{OLS}}, \quad \text{(F6)}$$

where $\tilde{X}_2 = QX_2$. The $N_h \times 1$ vector of holdout prediction errors for both IV and OLS are (where the subscript 'M' indicates either IV or OLS)

$$\hat{e}_{M\text{-pseudo},2} = \tilde{Y}_2 - \hat{Y}_{\text{M-pseudo},2} = \tilde{X}_2\beta + \tilde{\epsilon}_2 - \tilde{X}_2\hat{\beta}_M = \tilde{\epsilon}_2 + \tilde{X}_2\big(\beta - \hat{\beta}_M\big), \tag{F7}$$

and the sum of the squared prediction errors is

$$\hat{e}^{'}_{M\text{-pseudo},2}\hat{e}_{M\text{-pseudo},2} = \tilde{\epsilon}_2'\tilde{\epsilon}_2 + \tilde{\epsilon}_2'\tilde{X}_2(\beta - \beta_M) + (\beta - \beta_M)'\tilde{X}_2'\tilde{\epsilon}_2 + (\beta - \beta_M)'\tilde{X}_2'\tilde{X}_2(\beta - \hat{\beta}_M).$$

As the total sample tends to infinity (by growing both the estimation and holdout sample), the probability limit of $\tilde{X}_2^{'}\tilde{\epsilon}_2/\mathrm{N_h}$ is zero because X_2 is strictly exogenous with respect to ϵ_2. Hence, the 2nd and 3rd terms in the sum of the squared prediction errors drop in the limit (for the mean). For IV, we also have that $\hat{\beta}_{\text{IV}}$ is consistent for β, hence the last term also converges to zero, but not for OLS because $\hat{\beta}_{\text{OLS}}$ is inconsistent. Hence,

$$\begin{aligned} \text{A.MSE}_{\text{IV-pseudo}} &= \text{plim } \hat{e}^{'}_{IV\text{-pseudo},2}\hat{e}_{IV\text{-pseudo},2}/\,N_h \\ &= \text{plim } \tilde{\epsilon}_2'\tilde{\epsilon}_2/N_h = (T_2 - 1)/T_2\,\sigma_\epsilon^2, \end{aligned} \tag{F8}$$

because $E\big(\tilde{\epsilon}_{it,2}^2\big) = E\big(\epsilon_{it,2} - \bar{\epsilon}_{i,2}.\big)^2 = (T_2 - 1)/T_2\,\sigma_\epsilon^2$ where $\bar{\epsilon}_{i,2}. = 1/T_2 \sum_t \epsilon_{it,2}$ (e.g., Slutsky's theorems and Khintchine's theorem[10], Greene 2003, p. 113 and p. 365). Hence, the asymptotic mean squared prediction error for IV with pseudo exogenous regressors is generally smaller than σ_ϵ^2, however, the A.MSE may be multiplied by $T_2/(T_2 - 1)$ to obtain σ_ϵ^2. Similarly, for OLS

$$\begin{aligned} \text{A.MSE}_{\text{OLS-pseudo}} &= \text{plim } \hat{e}^{'}_{OLS\text{-pseudo},2}\hat{e}_{OLS\text{-pseudo},2}/\,N_h \\ &= \text{plim } \tilde{\epsilon}_2'\tilde{\epsilon}_2/N_h + \{\text{plim}(\epsilon_1'X_1/N_e)\text{plim}(X_1'X_1/N_e)^{-1}\times \\ &\quad \text{plim}\big(\tilde{X}_2'\tilde{X}_2/N_h\big)\text{plim}(X_1'X_1/N_e)^{-1}\text{plim}(X_1'\epsilon_1/N_e)\}. \end{aligned} \tag{F9}$$

[10] Khintchine's theorem is for i.i.d. random variables, which is not the case here because the observations within subject *i* are not independent. However, independence holds across the subjects *i=1,…,N*. Hence, we may view the set of *T* observations on the *i*th subject as a single observation and apply standard convergence arguments by growing *N* (e.g., Greene 2003, p. 366).

Here $\text{plim}(\tilde{X}_2'\tilde{X}_2/N_h)$ is the (asymptotic) variance-covariance matrix of the holdout regressors in deviations from individual means. This last term in $\text{A.MSE}_{\text{OLS-pseudo}}$ has no simple expression, but it is nonnegative (the product of the three (asymptotic) matrices is positive definite since all are of full rank, e.g., Davidson and MacKinnon 1993, pp. 787-788). We write for notational simplicity,

$$\delta^2 = \text{plim}(\epsilon_1'X_1/N_e)\text{plim}(X_1'X_1/N_e)^{-1}\text{plim}(\tilde{X}_2'\tilde{X}_2/N_h)\text{plim}(X_1'X_1/N_e)^{-1}\text{plim}(X_1'\epsilon_1/N_e) \geq 0$$

so that $\text{A.MSE}_{\text{OLS-pseudo}} = (T_2 - 1)/T_2\,\sigma_\epsilon^2 + \delta^2$.

Hence, $\text{A.MSE}_{\text{IV-pseudo}} \leq \text{A.MSE}_{\text{OLS-pseudo}}$. This result holds for holdout samples A, B, or C, because in each case the transformation Q eliminates the unobserved individual level effects α_i so that the transformed regressors are (pseudo-)exogenous.

G. Models with limited dependent variables

In a simulation study we analyze whether the main insights from our study extent to models with limited dependent variables. We use the following model to generate the data:

$$Y_i^* = X_i + \epsilon_i. \tag{G1}$$

$$X_i = Z_i + \psi_i, \tag{G2}$$

for $i = 1, \ldots, N$ where the errors (ϵ_i, ψ_i) have a joint normal distribution with mean 0 and variance 1. The covariance between the errors is $\sigma_{\epsilon\psi} = 0.93$ which is equivalent to the high extent of endogeneity case in the main document. The instrument explains 50% of the variance in X and we take $Z_i \sim N(0,1)$. The observed choices are related to the latent variable Y_i^* as follows: $Y_i = 1$ if $Y_i^* > 0$, and $Y_i = 0$ otherwise. We set the total sample size to N=2000 ($N = N_1 + N_2 = 1000 + 1000$) and generate 100 data sets. In each of the 100 datasets we estimate a standard probit model that treats X as exogenous by estimating (G1) only, and an IV-probit model that considers both (G1) and (G2) jointly. We assess the predictive performance by computing the mean squared difference between the observed choices Y_i and the predicted choice probability $\hat{p}\{Y_i = 1\}$ from the fitted models. The results are summarized in Table A.G1 (for an endogenous holdout sample) and Table A.G2 (for an exogenous holdout sample).

The results in Table A.G1 indicate that the main conclusions obtained in this study—that holdout sample validation does not uniquely identify the correct estimated relationship between X and Y when X is endogenous—also holds in case of limited dependent variables. Furthermore, when the holdout sample regressors are exogenous, the fitted IV probit model does produce the lowest holdout sample forecast errors (Table A.G2),

Table A.G1
Mean squared holdout prediction errors for probit and probit-IV for an *endogenous* holdout sample.

Information set	Probit	Probit-IV
Holdout regressors only (X)	.09	.11
Holdout regressors (X) and instruments (Z)	.04	.04

Table A.G2
Mean squared holdout prediction errors for probit and probit-IV for an *exogenous* holdout sample.

Information set	Probit	Probit-IV
Holdout regressors only (X)	.18	.17
Holdout regressors (X) and instruments (Z)	.23	.23

H. Heterogeneous regression coefficients

We investigate by means of a small simulation study whether the presence of heterogeneous regression coefficients affect the conclusions in section 2.2 of the main document. The following model was used to generate the data:

$$Y_{it} = \beta_{0i} + \beta_{1i} X_{it} + \epsilon_{it} \quad \text{(H1)}$$

$$X_{it} = \gamma_{0i} + \gamma_{1i} X_{it} + \psi_{it} \quad \text{(H2)}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. Here $(\beta_{0i}, \beta_{1i})' \sim N_2(\mu_\beta, \Omega_\beta)$, $(\gamma_{0i}, \gamma_{1i})' \sim N_2(\mu_\gamma, \Omega_\gamma)$, and $(\epsilon_{it}, \psi_{it}) \sim N_2(0, \Sigma)$. We use Bayesian estimation to obtain individual-level estimates for the heterogeneous regression effects (e.g., Allenby and Rossi, 1999). We adapted the fitted models in (4) and (5) from the main document and (D3) and (D6) for this heterogeneous case and we used the posterior means of the β_i's to generate the holdout predictions. Here, the holdout sample contains new observations in dimension t. If a separate regression model for the heterogeneous regression effects is specified, where the regression effects are explained by another set of (subject-level) regressors (e.g. Rossi et al. 2005, p. 132), then holdout sample predictions may be obtained in the dimension *i* as well. For simplicity, we did not consider that case here.

We specify conjugate, non-informative, priors for the means (μ_β and μ_γ) and variances (Ω_β, Ω_γ, and Σ). The true parameter values are $\mu_\beta = \mu_\gamma = (0,1)'$, $\Omega_\beta = \Omega_\gamma = \begin{bmatrix} 1 & 0 \\ 0 & 0.5 \end{bmatrix}$, and $\Sigma = \begin{bmatrix} 0.5 & 0.33 \\ 0.33 & 0.5 \end{bmatrix}$. We set $N = T = 100$, and generated 100 datasets in total. For each dataset, the Hierarchical Bayes (HB) model estimates (H1), and the Hierarchical Bayes Instrumental Variables (HB-IV) model estimates both (H1) and (H2) jointly.

The results (averaged across the 100 datasets) are given in Table A.H1 (for an endogenous holdout sample) and Table A.H2 (for an exogenous holdout sample). It follows that the main conclusions of this study are not affected by the presence of heterogeneity in the regression coefficients. That is, for an endogenous holdout sample, the non-corrected model predicts better than the endogeneity-corrected model (unless both use *Z*, in which case the predictions are the same).

Furthermore, when the holdout sample regressors are *exogenous*, the approach that corrects for endogeneity and heterogeneity and only uses *X* to predict gives the lowest mean-squared prediction error in the holdout sample.

Table A.H1
Mean squared holdout prediction errors for HB and HB-IV in case of heterogeneous responses for an *endogenous* holdout sample.

Information set	HB	HB-IV
Holdout regressors only (X)	0.42	0.51
Holdout regressors (X) and instruments (Z)	0.29	0.29

Table A.H2
Mean squared holdout prediction errors for HB and HB-IV in case of heterogeneous responses for an *exogenous* holdout sample.

Information set	HB	HB-IV
Holdout regressors only (X)	0.60	0.51
Holdout regressors (X) and instruments (Z)	0.72	0.72

I. References

Ailawadi, K.L., P.K. Kopalle, S.A. Neslin. 2005. Predicting Competitive Response to a Major Policy Change: Combining Game-Theoretic and Empirical Analyses. *Marketing Science*. **24**(1) 12-24.

Allenby, G., P. Rossi. 1999. Marketing models of heterogeneity. *Journal of Econometrics*. **89**(1/2) 57-78.

Andrews, R.L., I.S. Currim. 2009. Multi-stage purchase decision models: Accommodating response heterogeneity, common demand shocks, and endogeneity using disaggregate data. *International Journal of Research in Marketing*. **26**(3) 197-206.

Bass, F.M., N. Bruce, S. Majumdar, B.P.S. Murthi. 2007. Wearout Effects of Different Advertising Themes: A Dynamic Bayesian Model of the Advertising-Sales Relationship. *Marketing Science*. **26**(2) 179-195.

Besanko, D., S. Gupta, D. Jain. 1998. Logit Demand Estimation under Competitive Pricing Behavior: An Equilibrium Framework. *Management Science*. **44**(11) 1533-1547.

Briesch, R.A., W.R. Dillon, R.C. Blattberg. 2008. Treating Zero Brand Sales Observations in Choice Model Estimation: Consequences and Potential Remedies. *Journal of Marketing Research (JMR)*. **45**(5) 618-632.

Chan, T.Y., V. Padmanabhan, P.B. Seetharaman. 2007. An Econometric Model of Location and Pricing in the Gasoline Market. *Journal of Marketing Research*. **44**(4) 622-635.

Chintagunta, P.K. 2001. Endogeneity and Heterogeneity in a Probit Demand Model: Estimation Using Aggregate Data. *Marketing Science*. **20**(4) 442-456.

Chintagunta, P.K., J.-P. Dubé. 2005. Estimating a Stockkeeping-Unit-Level Brand Choice Model That Combines Household Panel Data and Store Data. *Journal of Marketing Research*. **42**(3) 368-379.

Danaher, P.J., A. Bonfrer, S. Dhar. 2008. The Effect of Competitive Advertising Interference on Sales for Packaged Goods. *Journal of Marketing Research*. **45**(2) 211-225.

Davidson, R., J. G. MacKinnon. 1993. *Estimation and Inference in Econometrics*. Oxford: Oxford University Press.

Dong, X., P. Manchanda, P.K. Chintagunta. 2009. Quantifying the Benefits of Individual-Level Targeting in the Presence of Firm Strategic Behavior. *Journal of Marketing Research*. **46**(2) 207-221.

Drèze, X., P. Nisol, N. J. Vilcassim. 2004. Do Promotions Increase Store Expenditures? A Descriptive Study of Household Shopping Behavior. *Quantitative Marketing and Economics*. **2**(1) 59-92.

Ferguson, T. S. 1996. *A Course in Large Sample Theory*. London: Chapman & Hall.

Fischer, M., S. Albers. 2010. Patient- or Physician-Oriented Marketing: What Drives Primary Demand for Prescription Drugs? *Journal of Marketing Research*. **47**(1) 103-121.

Greene, W. H. 2003. *Econometric Analysis*, Upper Saddle River, NJ: Prentice Hall.

Kleibergen, F., E. Zivot. 2003. Bayesian and classical approaches to instrumental variables regression. *Journal of Econometrics.* **114**(1) 29-72

Kumar, V., J. Fan, R. Gulati, P. Venkat. 2009. Marketing-Mix Recommendations to Manage Value Growth at P&G Asia-Pacific. *Marketing Science*. **28**(4) 645-655.

Kumar, V., R. Venkatesan, W. Reinartz. 2008. Performance Implications of Adopting a Customer-Focused Sales Campaign. *Journal of Marketing*. **72**(5) 50-68.

Leenheer, J., H.J. van Heerde, T.H.A. Bijmolt, A. Smidts. 2007. Do loyalty programs really enhance behavioral loyalty? An empirical analysis accounting for self-selecting members. *International Journal of Research in Marketing*. **24**(1) 31-47.
Li, S., K. Srinivasan, B. Sun. 2009. Internet Auction Features as Quality Signals. *Journal of Marketing*. **73**(1) 75-92.
Manchanda, P., P.E. Rossi, P.K. Chintagunta. 2004. Response Modeling with Nonrandom Marketing-Mix Variables. *Journal of Marketing Research*. **41**(4) 467-478.
Narayanan, S., R. Desiraju, P.K. Chintagunta. 2004. Return on Investment Implications for Pharmaceutical Promotional Expenditures: The Role of Marketing-Mix Interactions. *Journal of Marketing*. **68**(4) 90-105.
Neslin, S.A. 1990. A Market Response Model for Coupon Promotions. *Marketing Science*. **9**(2) 125-145.
Petrin, A., K. Train. 2010. A Control Function Approach to Endogeneity in Consumer Choice Models. *Journal of Marketing Research*. **47**(1) 3-13.
Richards, T. J. 2007. A Nested Logit Model of Strategic Promotion. *Quantitative Marketing and Economics*. **5** 63-91.
Rossi, P. E., G. M. Allenby, and R. McCulloch. 2005. *Bayesian Statistics and Marketing*. John Wiley&Sons, Chichester.
Song, I., P.K. Chintagunta. 2006. Measuring Cross-Category Price Effects with Aggregate Store Data. *Management Science*. **52**(10) 1594-1609.
Van Dijk, A., H. J. van Heerde, P. S. H. Leeflang, D. R. Wittink. 2004. Similarity-Based Spatial Methods for Estimating Shelf Space Elasticities from Correlational data. *Quantitative Marketing & Economics*. 2(3) 257-277.
Van Nierop, E., D. Fok, P.H. Franses. 2008. Interaction between Shelf Layout and Marketing Effectiveness and Its Impact on Optimizing Shelf Arrangements. *Marketing Science*. **27**(6) 1065-1082.
Venkatesan, R., V. Kumar. 2004. A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy. *Journal of Marketing*. **68**(4) 106-125.
Venkatesan, R., V. Kumar, T. Bohling. 2007. Optimal Customer Relationship Management Using Bayesian Decision Theory: An Application for Customer Selection. *Journal of Marketing Research.* **44**(4) 579-594.
Verbeek, M. 2008. *A Guide to Modern Econometrics*. John Wiley & Sons, Hoboken, NJ.
Zhu, T., V. Singh, M.D. Manuszak. 2009. Market Structure and Competition in the Retail Discount Industry. *Journal of Marketing Research*. **46**(4) 453-466.