Political Analysis (2012) 20:369–386
doi:10.1093/pan/mps020

A Mixture Model for Middle Category Inflation in Ordered Survey Responses

Benjamin E. Bagozzi
Department of Political Science, Penn State University, University Park, PA 16802
e-mail: beb196@psu.edu (corresponding author)

Bumba Mukherjee
Department of Political Science, Penn State University, University Park, PA 16802
e-mail: sxm73@psu.edu

Edited by R. Michael Alvarez

Recent research finds that, for social desirability reasons, uninformed individuals disproportionately give "neither agree nor disagree" type responses to survey attitude questions, even when a "do not know" option is available. Such "face-saving" responses inflate the indifference (i.e., middle) categories of ordered attitude variables with nonordered responses. When such inflation occurs within the middle category of one's ordered dependent variable, estimates from ordered probit (and ordered logit) models are likely to be unreliable and inefficient. This article develops a set of mixture models that estimate and account for the presence of "face-saving" responses in middle categories of ordered survey response variables, and applies these models to (1) simulated data and (2) a commonly studied survey question measuring attitudes toward European Union (EU) membership among individuals in EU-candidate countries. Results from the survey data set and the Monte Carlo experiments suggest that, when middle category inflation is present in one's ordered dependent variable, the estimates obtained from middle category mixture models are less biased than—and in some cases substantively distinct from—the estimates obtained from "naive" ordered probit models.

Survey questions of the attitude or opinion variety—where respondents are asked to place themselves along an ordered categorical scale that ranges from negative to positive responses—often include a middle category response to accommodate attitudes of "indifference" or "neutrality."[1] These respondents may be fully informed or uninformed (or partially informed) about the political issues captured in such survey questions (Delli Carpini and Keeter 1997; Mondak 2000; Kuklinski et al. 2000; Barabas 2002). Problematically, studies show that uninformed respondents often desist from choosing "do not know" responses in these settings (Fiske, Lau, and Smith 1990; Zaller 1992; Mondak 2000; Mondak and Creel Davis 2001). Instead, many politically uninformed individuals consistently take positions of indifference on political issues when middle category responses such as "neither agree nor disagree" exist (Ferber 1956; Alvarez and Franklin 1994a; Sturgis, Roberts, and Smith 2010).[2] Doing so allows uninformed respondents to "save face" by *not* revealing their lack of knowledge about key political issues or to appear highly informed, which is socially desirable.[3] But this practice also conflates ordered middle category responses with the inherently

Authors' note: An earlier version of this article was presented as a poster at the 2011 Political Methodology Meeting. The authors wish to acknowledge the valuable suggestions that they received from the editor and reviewers of *Political Analysis*, Chris Zorn, Phil Schrodt, Will H. Moore, Daniel W. Hill, Eric Plutzer, John Freeman, Robert J. Franzese Jr., and the 2011 Political Methodology meeting attendees. This work was supported by the Pennsylvania State University's College of Liberal Arts under the Forrest Crawford Graduate Scholarship. Replication materials can be found at Bagozzi and Mukherjee (2012). Supplementary Materials for this article are available on the *Political Analysis* Web site.

[1]Middle category responses include, e.g., "neither agree nor disagree" or "neither good nor bad."
[2]Studies also show that respondents who are more uncertain about their position on certain policy issues are more likely to opt for the middle category response on survey questions that track their perception about their senator's position on such issues (Bradey and Sniderman 1991; Alvarez and Franklin 1994a,b).
[3]Zaller (1992); Mondak (2000); Burden (2000); Berinsky (2002).

unordered position of “do not know.” As discussed below, employing ordered survey–response-dependent variables in which uninformed individuals opt for the middle category to save face generates serious econometric challenges for researchers. This article builds on existing survey–response mixture models (Jackson 1993; Harris and Zhao 2007) to propose a new mixture model (the middle-inflated ordered probit model) that addresses these challenges.

To understand the problem discussed above, consider extant studies of public attitudes toward European Union (EU) membership that analyze a Eurobarometer survey question that asks respondents whether their country’s membership in the EU is bad, neither good nor bad, or good (Gabel 1998; Carey 2002; Elgün and Tillman 2007). Responses to this question are measured on an ordered scale that ranges from “a bad thing” to a “good thing” with the middle category of “neither good nor bad.” We show in this article’s empirical section that the middle category of “neither good nor bad” contains responses from informed respondents who have substantial knowledge about the impact of EU membership *and* from a significant majority of *uninformed* individuals who save face by opting for this middle category despite their lack of knowledge about the EU. This finding is not surprising given that most uninformed respondents in Europe and in the United States often place themselves in the “neither/nor” (i.e., middle) category of ordered survey response measures that track their attitudes or opinions about political issues (Ferber 1956; Alvarez and Franklin 1994a; Sturgis, Roberts, and Smith 2010).

The example discussed above suggests that middle category responses to ordered survey variables typically emerge from two distinct sources: (1) informed respondents who are truly indifferent and (2) uninformed respondents who choose “neither/nor” to save face. When uninformed respondents opt for “neither/nor” over “do not know,” the middle category of ordered survey responses becomes *inflated* with unordered “face-saving” responses. Middle category inflation of this sort poses methodological challenges for scholars who analyze such measures as dependent variables in ordered probit (OP) or ordered logit (OL) models. For instance, unlike true indifference-responses, middle-inflated responses do not represent the actual midpoint of an underlying (directional) preference dimension, but rather correspond to *nondirectional* positions that a respondent is unable to place along the ordered dimension scale (Sturgis, Roberts, and Smith 2010). As a result, middle category “face-saving” responses add measurement error to one’s dependent variable and violate assumptions of ordinality, which reduces the efficiency and consistency of one’s estimates (Bound, Brown, and Mathiowetz 2001). Finally, if a given covariate affects *both* the inflation process producing “face-saving” responses *and* the ordered outcome of interest, OP (or OL) models will incorrectly estimate that variable’s direct effect through their (mis)attribution of inflation-stage effects to the ordered outcome process.

To address the aforementioned problems, we build on existing survey–response mixture models (Jackson 1993; Harris and Zhao 2007) to develop the middle-inflated ordered probit (MiOP) model that statistically accounts for inflation in the middle category of an ordered-dependent variable. It does so by explicitly modeling the potential for a dual data-generating process (d.g.p.) within the middle category of one’s ordered-dependent variable. Specifically, the MiOP model assumes that a given ordered variable is composed of two distinct latent processes: (1) a probit “inflation” model and (2) an OP “outcome” model; and then estimates both processes under a single system. Different covariates can be included within each of the proposed MiOP model’s two latent stages and, moreover, correlated disturbances can be estimated between the two latent equations of the MiOP model (a variant hereafter referred to as the MiOPC model). A series of Monte Carlo (MC) experiments and an application to a data set on attitudes toward EU membership presented below suggest that the MiOP(C) models outperform OP models in recovering one’s true “outcome stage” estimates when the middle category of the ordered-dependent variable is related to two distinct d.g.p’s.

This study proceeds as follows. In the next section, we present the MiOP model with and without correlated errors, and compare the MiOP(C) models to zero-inflated estimators, Heckman-type selection models, and multinomial survey–response mixture models. This is followed by a discussion of the results derived from Monte Carlo simulations that evaluate the performance of the OP, MiOP, and MiOPC models when the middle category of the ordered-dependent variable is inflated. We then report the estimates obtained from applying the OP, MiOP, and MiOPC models to a survey data set on individual attitudes toward EU membership, and perform a battery of additional tests to assess the robustness of our MiOP(C) model results. We conclude by suggesting that the MiOP(C) model can be applied to a variety of additional survey studies, such as those pertaining to immigration support (e.g., Hainmueller and Hiscox 2010; Richey 2010), future economic assessment (e.g., Gerber and Huber 2010), and political ideology (Gerber et al. 2010). We also suggest in the conclusion that extensions of the model presented here may potentially have implications for methodological research on (1) heterogeneity in survey responses about vote choice and attitudes (e.g., Arceneaux and Kolodny 2009; Alvarez and Franklin 1994a, 1994b) and (2) misreporting by respondents in the context of voter turnout (e.g., Katz and Katz 2010).

1 The MiOP and MiOPC models

The MiOP model is a mixture model that addresses the issue of middle category inflation in ordered survey–response-dependent variables by statistically accounting for two groups of respondents (informed and uninformed face-saving respondents) who opt for the "neither/nor" response. It does so by combining the following two latent equations: a "split" probit equation in the first stage (estimating the effect of covariates on the probability with which respondents come from the uninformed versus informed group) and an ordered probit equation in the outcome (second) stage (estimating the effect of a second set of covariates on the probabilities of observing each ordered survey response category, conditional on respondents being in the informed group). Each of these two latent equations contains a stochastic error term that may be correlated or independent. We thus first present below the MiOP model, which assumes that the errors from the two latent equations are independent of each other, and then the MiOPC model, which assumes that the two error terms are correlated with each other.

To begin with, suppose that we have an ordered survey–response-dependent variable y_i where $i \in \{1, 2, \ldots N\}$. Suppose further that y_i is observable and assumes the discrete ordered values of $j = 0, 1, 2, \ldots J$. Let s_i denote a binary variable that indicates a split between regime 0 ($s_i = 0$) and regime 1 where $s_i = 1$. In the context of the survey example described earlier, the observations in regime 0 ($s_i = 0$) include respondents who belong to the uninformed group, while observations in regime 1 ($s_i = 1$) include survey respondents who are in the informed group. Note that s_i is related to the latent-dependent variable s_i^* such that $s_i = 1$ for $s_i^* > 0$ and $s_i = 0$ for $s_i^* \leq 0$. The latent variable s_i^* represents the propensity with which respondents enter regime 1 (i.e., are informed) and is given by the following split probit "inflation" equation:

$$s_i^* = \mathbf{z}_i'\gamma + u_i. \tag{1}$$

In equation (1), $\mathbf{z}_i'$ is the vector of covariates, γ is the vector of coefficients, and u_i is a standard-normal distributed error term. Hence, the probability of respondent i being in regime 1 is $\Pr(s_i = 1|\mathbf{z}_i) = \Pr(s_i^* > 0|\mathbf{z}_i) = \Phi(\mathbf{z}_i'\gamma)$, and the probability that respondent i is in regime 0 is $\Pr(s_i = 0|\mathbf{z}_i) = \Pr(s_i^* \leq 0|\mathbf{z}_i) = 1 - \Phi(\mathbf{z}_i'\gamma)$, where $\Phi(.)$ is the standard normal c.d.f.

The outcome equation of the MiOP(C) model is developed from the ordered probit equation that is defined as

$$\widetilde{y}_i^* = \mathbf{x}_i'\beta + \varepsilon_i \tag{2}$$

$$\widetilde{y}_i = \left\{ \begin{array}{l} 0 \text{ if } \widetilde{y}_i^* \leq 0 \\ j \text{ if } \mu_{j-1} < \widetilde{y}_i^* \leq \mu_j (j = 1, \ldots, J-1), \\ J \text{ if } \mu_{J-1} \leq \widetilde{y}_i^* \end{array} \right\} \tag{3}$$

where $\mathbf{x}_i'$ is a vector of covariates, β is the vector of coefficients, ε_i is a standard normal distributed error term, and $j = 1, 2, \ldots J-1$. μ_j is the vector of boundary parameters that need to be estimated in addition to β. We assume throughout, without loss of generality, that $\mu_{j=0} = 0$. If the error terms from the split probit equation (u_i) and the ordered probit equation (ε_i) are *not* correlated, then the augmented ordered probit outcome equation of the MiOP model is defined as

$$\Pr(y_i) = \left\{ \begin{array}{l} \Pr(y_i = 0|\ \mathbf{x}_i, \mathbf{z}_i) = [\Phi(\mathbf{z}_i'\gamma)\Phi(-\mathbf{x}_i'\beta)] \\ \Pr(y_i = j|\ \mathbf{x}_i, \mathbf{z}_i) = [1 - \Phi(\mathbf{z}_i'\gamma)] + \Phi(\mathbf{z}_i'\gamma)[\Phi(\mu_j - \mathbf{x}_i'\beta) - \Phi(\mu_{j-1} - \mathbf{x}_i'\beta)]\ (j = 1, \ldots, J-1) \\ \Pr(y_i = J|\ \mathbf{x}_i, \mathbf{z}_i) = \Phi(\mathbf{z}_i'\gamma)[1 - \Phi(\mu_{J-1} - \mathbf{x}_i'\beta)] \end{array} \right\}. \tag{4}$$

The expression in (4) provides the full probabilities of the augmented ordered probit (outcome) equation of the MiOP model; we label these probabilities as outcome probabilities for convenience. Put together, then, the split probit equation in (1) constitutes the first stage of the MiOP model, while the augmented ordered probit outcome equation in (4) constitutes the second stage of the MiOP model.

Now suppose that the error terms in the two separate latent equations of the MiOP model (u_i and ε_i) are correlated, as they correspond to the same unit of analysis. If the error terms u_i and ε_i are correlated and follow a bivariate normal distribution with correlation coefficient $\rho_{\varepsilon u}$, then the augmented ordered probit outcome equation of the MiOPC model is

$$\Pr(y_i) = \left\{ \begin{array}{l} \Pr(y_i = 0|\ \mathbf{x}_i, \mathbf{z}_i) = \Phi_2(\mathbf{z}_i'\gamma, -\mathbf{x}_i'\beta; -\rho_{\varepsilon u}) \\ \Pr(y_i = j|\ \mathbf{x}_i, \mathbf{z}_i) = [1 - \Phi(\mathbf{x}_i'\beta)] + \left\{ \begin{array}{l} \Phi_2(-\mathbf{z}_i'\gamma, \mu_{j-1} - \mathbf{x}_i'\beta; -\rho_{\varepsilon u}) \\ -\Phi_2(-\mathbf{z}_i'\gamma, -\mathbf{x}_i'\beta; -\rho_{\varepsilon u}) \end{array} \right\} \\ \Pr(y_i = J|\ \mathbf{x}_i, \mathbf{z}_i) = \Phi_2(\mathbf{z}_i'\gamma, \mathbf{x}_i'\beta - \mu_{J-1}; -\rho_{\varepsilon u}), \end{array} \right\} \tag{5}$$

where $\Phi_2(.)$ denotes the c.d.f. of the standardized bivariate normal distribution. The expression in equation (5) provides the full outcome probabilities of the augmented ordered probit equation—the second stage—of the MiOPC model, while the split probit equation in (1) constitutes the first stage of the MiOPC model.

As mentioned earlier, the MiOP and MiOPC models presented here jointly estimate the split probit equation and the relevant augmented OP outcome equation. More importantly, note that the probability of the middle category in the augmented ordered probit equation of the MiOP model (4) and MiOPC model (5) is modeled conditional upon the probability of an observation being assigned a middle category value in the ordered probit process plus the probability of it being in regime 0 (the *uninformed* group) from the split probit (i.e., inflation) equation. This feature helps researchers to account for middle category inflation in ordered survey–response-dependent variables that is partly engendered by a high proportion of uninformed respondents who opt the "neither/nor" (middle category) response to save face. It also helps researchers to take into account that the inflated middle category of ordered survey response variables contain responses from the two distinct groups mentioned earlier.

Given these features, our MiOP(C) models are directly analogous to existing inflated estimators such as the zero inflated (i) ordered probit (ZiOP) model with and without correlated errors and (ii) Poisson (ZiP) models. When applied to survey–response data, inflated models of these sorts can also be interpreted as "self selection models"; as they exhibit a number of similarities with two-stage selection models (Heckman 1979; Heckman and Sedlacek 1990; Winkelman 1998). Indeed, inflation processes are statistically and conceptually similar to selection processes in that each process produces an undesirable sample of outcome observations, which must then be "corrected for" via (1) explicit estimation of the inflation/selection process and (2) an incorporation of the resultant inflation/selection probabilities into one's outcome stage estimation. The key difference between these two processes, however, is that binary selection processes truncate desirable observations from one's outcome sample, whereas binary inflation processes augment outcome samples with undesirable observations. Hence, inflated-estimators require that *all* observations appear within *both* estimation stages while selection models necessitate that *some* (selection stage) observations do not occur in one's outcome stage. The MiOPC model therefore offers two distinct advantages over a Heckman OP selection model. First, it enables one to correctly estimate ordered outcomes when sample bias arises within intermediate, rather than initial categories. Second, the MiOPC model allows one to account for the "self-selection" biases that arise when heterogeneous survey respondents select into—rather than out-of—one's sample of interest.

The MiOP and MiOPC models are also comparable—in both theoretical underpinning and statistical approach—to the multinomial survey–response mixture model proposed by Jackson (1993). Jackson's model addresses the survey–response challenge of "response-guessing" by uninformed respondents, a multiresponse category inflation process. Similar to our argument above, Jackson explicitly assumes that observed survey–responses are a mixture of two latent processes—one relating to the propensity of a respondent being "informed" and one relating to a respondent's true preferences—which together determine individuals' actual question responses. The author then uses a (categorical) variation of the latent multiindicator, multicause (MIMC) model to estimate both processes' contributions to the observed survey responses. Hence, both Jackson's model and the MiOP(C) models allow researchers to (1) estimate the effects of particular covariates on the likelihood of a respondent being informed and (2) assess the extent to which survey-question responses depend on a respondent being informed[4]—albeit for different types of discrete-dependent variables and with different assumptions of uninformed response-set behaviors. Therefore, although the MiOPC model improves upon Jackson's model through its allowance for correlated disturbances, the similarities between Jackson's model and our own suggest that many of the fragility concerns raised by Jackson (1993, 43) may also apply to the MiOP(C) models, and it is to these concerns that we now turn.

As illustrated above, inflated models share a number of similarities with selection models and other related multiequation estimation techniques. This is especially the case among inflated models that allow for correlated disturbances (e.g., the MiOPC), which in some instances are in fact identical to bivariate estimators with partial observability (Xiang 2010). The similarities between inflated models and selection-type models suggest that careful attention must be paid to issues of exclusion restriction and model identification when using the former. Although researchers continue to debate the relevance of exclusion restrictions in the context of uncorrelated inflated models (Winkelman 1998; Harris and Zhao 2007; Burger, Oort, and Linders 2009, 176), numerous scholars have rightly noted that exclusion restrictions are nevertheless required for the proper identification and estimation of many latent variable mixture models (Jackson 1993), and especially for limited-dependent variable mixture models that incorporate correlated errors (Xiang 2010). We therefore explore this issue below in our Monte Carlo analysis of the MiOP(C) model and in the application of this model to a survey response on support for EU membership.

Having described above the MiOP and MiOPC models, we turn to define the (log-)likelihood function for these two models. Specifically, let $\theta = (\gamma', \beta', \mu')'$ for the full MiOP model and let $\widehat{\theta} = (\gamma', \beta', \mu', \rho_{\varepsilon u})'$ for the full MiOPC model. The likelihood of the MiOP model for an i.i.d. sample of $i \in \{1, 2, \ldots N\}$ observations is $\mathcal{L}(\theta) = \prod_{i=1}^{N} \prod_{j=0}^{J} [\Pr(y_i = j | \mathbf{x}_i, \mathbf{z}_i, \theta)]^{d_{ij}}$, which is fully defined by the following:

$$
\begin{aligned}
\mathcal{L}(\theta) = & \prod_{i=1}^{N} \prod_{j=0}^{m-1} [\Pr(s_i = 1) \Pr(\widetilde{y}_i = j)]^{d_{ij}} \\
& \times \prod_{i=1}^{N} \prod_{j=m}^{m} [\Pr(s_i = 0) + \Pr(s_i = 1) \Pr(\widetilde{y}_i = j)]^{d_{ij}} \\
& \times \prod_{i=1}^{N} \prod_{j>m}^{J} [\Pr(s_i = 1) \Pr(\widetilde{y}_i = j)]^{d_{ij}},
\end{aligned} \tag{6}
$$

[4]As argued above, feature (2) in turn enables researchers to obtain unbiased estimates of their covariates' direct effects on the dependent variable of interest.

where m is the middle category of ordered-dependent variable y and where $d_{ij}=1$ if individual i chooses category j, or is $d_{ij}=0$ otherwise. The likelihood function for the correlated (i.e., MiOPC) model is $L(\widehat{\theta})=\prod_{i=1}^{N}\prod_{j=0}^{J}[\Pr(y_i=j|\mathbf{x}_i,\mathbf{z}_i,\widehat{\theta})]^{d_{ij}}$, which is fully defined as

$$
\begin{aligned}
\mathcal{L}(\widehat{\theta}) = & \prod_{i=1}^{N}\prod_{j=0}^{m-1}[\Pr(s_i=1,\widetilde{y}_i=j)]^{d_{ij}} \\
& \times \prod_{i=1}^{N}\prod_{j=m}^{m}[\Pr(s_i=0)+\Pr(s_i=1,\widetilde{y}_i=j)]^{d_{ij}} \\
& \times \prod_{i=1}^{N}\prod_{j>m}^{J}[\Pr(s_i=1,\widetilde{y}_i=j)]^{d_{ij}}.
\end{aligned} \tag{7}
$$

From equation (6), the log likelihood function of the MiOP model can then be defined as $\ell(\theta)=\sum_{i=1}^{N}\sum_{j=0}^{J}d_{ij}\ln[\Pr(y_i=j|\mathbf{x}_i,\mathbf{z}_i,\theta)]$, where the outcome probabilities are given by equation (4). The log likelihood function of the MiOPC model is $\ell(\widehat{\theta})=\sum_{i=1}^{N}\sum_{j=0}^{J}d_{ij}\ln[\Pr(y_i=j|\mathbf{x}_i,\mathbf{z}_i,\widehat{\theta})]$, where the outcome probabilities are given by equation (5). The log-likelihood functions of the MiOP and MiOPC model can be consistently and efficiently estimated using maximum likelihood, which yields asymptotically normally distributed maximum likelihood estimates.[5]

2 Monte Carlo Experiments

We conduct three main Monte Carlo exercises to assess the performance of the OP, MiOP, and MiOPC models when the middle category of an ordered-dependent variable is "inflated" and thus generated from two distinct d.g.p's. The results (including tables and figures) from these experiments are reported and discussed in detail in the supplementary appendix to our article. Therefore, we simply summarize our findings here. For the first Monte Carlo exercise, we compare the performance of our OP, MiOP, and MiOPC models when the degree of inflation in the middle category of the ordered-dependent variable is set at a relatively conservative level of 30% and the number of observations varies (at Ns of 2000, 4000, and 8000), under both an MiOP and MiOPC d.g.p. We find here that (1) our MiOP and MiOPC models perform equally well (and decidedly better than the OP model) in recovering true values of interest when the d.g.p. is MiOP with 30% inflation and (2) our MiOPC model outperforms the MiOP model (which in turn outperforms the OP model) when the d.g.p. is MiOPC with 30% inflation, and $N>2000$.[6] These results suggest that if the researcher suspects there to be at least moderate levels of middle category inflation, MiOP and MiOPC models should be favored over OP models when one's data contain at least 2000 observations.

For the second Monte Carlo exercise, we hold the number of observations (i.e., N) fixed at 2000 and then explore how the OP, MiOP, and MiOPC models perform when the percent of middle category inflation in the ordered (survey–response)-dependent variable is increased above 30% (to 60% and then 90%), again under both an MiOP and MiOPC d.g.p. Our findings here indicate that (1) increasing the proportion of middle category inflation above 30% consistently improves our MiOP(C) estimates but dramatically worsens our OP estimates, (2) our MiOP and MiOPC estimates remain comparable under an MiOP d.g.p., no matter the level of inflation, and (3) the MiOPC generally outperforms both the MiOP and OP models when inflation is greater than 30%.[7] Therefore, if one suspects there to be middle category inflation greater than 30%, then the MiOP and MiOPC models should be strongly favored over OP models.

Our final Monte Carlo exercise uses a simulated data set with 60% middle category inflation ($N=2.000$) to determine whether one needs to maintain exclusion restrictions for proper MiOP(C)

[5]Newton numerical optimization methods can be used to estimate the MiOP and MiOPC models. The first author of this article has written code to permit users to estimate the MiOP(C) model using *R*.
[6]See Table A.1 and Fig. A.1 in the Supplementary Appendix.
[7]See Fig. A.2 in the Supplementary Appendix.

estimation, again when our d.g.p's are either MiOP or MiOPC. To do so, we compare the OP, MiOP, and MiOPC estimates that are obtained when an exclusion restriction is incorporated into the MiOP(C) models and when it is not. For the MiOP model, we find that our results remain essentially unchanged no matter whether an exclusion restriction is maintained or not, and no matter whether the d.g.p. is MiOP or MiOPC. On the other hand, while our MiOPC models continue to outperform OP estimates under both d.g.p's, we find that the MiOPC model performs notably worse in recovering our true estimates—particularly in the outcome stages and for the correlation parameter ρ—when the exclusion restriction is ignored.[8] These Monte Carlo findings parallel those reported by Harris and Zhao 2007, 1084), and suggest that exclusion restrictions are needed for unbiased estimation of the MiOPC model. While it is difficult to ascertain whether the observed biases in our MiOPC findings here are due to underidentification or to increased collinearity (or both); it is likely that real-world data sets will be relatively more susceptible to both challenges. Thus, we can conclude here that exclusion restrictions will often be helpful—and in some cases critical—to accurate and efficient MiOP(C) model estimation, most notably for the MiOPC case.

3 Application to data

We now turn to applying the OP, MiOP, and MiOPC models to a commonly studied survey question mentioned earlier that is drawn from the 2002.2 Eurobarometer survey. This question (described below) focuses on citizens' attitudes toward membership in the European Union (EU) in 13 Central and Eastern European (CEE) candidate countries and has been analyzed for CEE as well as other EU countries (Gabel 1998; Carey 2002; Nelsen and Guth 2000; Tucker, Pacek, and Berinsky 2002; Hooghe and Marks 2005; Christin 2005; Elgün and Tillman 2007). We show below that the middle category of the ordered categorical response to this survey question on attitudes toward EU membership in CEE countries is "inflated." Thus, applying the OP, MiOP, and MiOPC model to the data generated by responses to this survey question permits us to assess whether our Monte Carlo results also hold in real-world data sets. The MiOP(C) coefficient estimates reported below also provide concrete information about the proportion of uninformed face-saving respondents in the Eurobarometer survey and their attitudes toward EU membership. Finally, unlike the OP model, the marginal effects of covariates from the MiOP(C) models provide a more accurate assessment of the relationship between the variables of interest and the outcome probabilities.

3.1 *The Eurobarometer Survey Data: Background*

Motivated by the literature on public support for EU membership,[9] a recent study by Elgün and Tillman (2007) uses ordered categorical responses to the following question in the Candidate Countries Eurobarometer 2002.2 survey to evaluate public attitudes toward EU membership in 13 CEE candidate countries: "Generally speaking, do you think that (your country's) membership of the European Union would be a good thing, a bad thing, or neither good nor bad?" Based on responses to this question, the discrete ordered-dependent variable—usually labeled as *EU support*—in Elgün and Tillman's (2007) study and in similar related studies[10] is coded as 1 for "a bad thing," 2 for "neither good nor bad," and 3 for "a good thing."

Respondents can also provide a "do not know" response to the question posited above. Elgün and Tillman (2007) deal with the relatively small share of "do not know" responses by adding these responses—ex post—to the middle "neither good nor bad" category of their ordinal-dependent survey–response variable in an effort to avoid dropping "do not know" responses altogether. This approach is fairly common among survey researchers studying EU-survey response questions with midpoint indifference categories (e.g., Rohrschneider 1990; Carey 2002; Rohrschneider 2002; Elgün and Tillman 2007), and likely exacerbates the problems of middle category inflation discussed

[8]See Table A.2 in the Supplementary Appendix.
[9]Gabel (1998); Carey (2002); Tucker, Pacek, and Berinsky (2002); Christin (2005).
[10]Carey (2002); Hooghe and Marks (2005); Christin (2005).

above. Because our objective is to directly address these survey studies in our empirical analysis, we accordingly follow their approach in our replication below, and add "do not know" responses to our middle category.[11]

Scholars then typically estimate ordered probit (or ordered logit) models to evaluate the effect of different covariates on *EU support* when testing theories of EU membership support (Carey 2002; Christin 2005; Elgün and Tillman 2007). As this is a methods exercise, we focus on three main theoretical claims (to save space) that are evaluated in extant research. The first explanation suggests that individuals with higher levels of cognitive mobilization, which results from increased political communication about European integration among peers, are more likely to have a positive attitude toward EU membership (Inglehart 1970; Gabel 1998; Karp, Banducci, and Bowler 2003). This claim is assessed by using an ordinal measure called *discuss politics*, which is coded as 1 if the respondent reports discussing politics with friends "never," as 2 if "occasionally," and as 3 if "frequently" (Gabel 1998; Elgün and Tillman 2007, 395). Second, scholars include *income* (measured in quartiles) in models of *EU support* to test the hypothesis that individuals with higher incomes are more likely to view EU membership as a "good thing" since they benefit from European integration (Gabel 1998; Carey 2002; Hooghe and Marks 2005; Elgün and Tillman 2007). Third, a dummy variable coded as 1 for *female* is often incorporated in the model by researchers on the premise that women are less likely to support EU membership as they are more vulnerable to the costs of integration that occur when states join the EU (Gabel 1998; Nelsen and Guth 2000; Carey 2002). Finally, the studies mentioned above add various controls to the OP specification that account for the respondents' *age*, occupational status, whether or not they are *unemployed*, their location (*rural*), whether or not they are educated at the college level (*college education*), and the extent to which they trust domestic political institutions (*political trust*) and are xenophobic.

The results reported in published research indicate consistent statistical support for the theoretical predictions summarized above (Gabel 1998; Carey 2002; Tucker, Pacek, and Berinsky 2002; Elgün and Tillman 2007). While these results are insightful, we find that the middle category response of "neither good nor bad" in the ordered *EU support*-dependent variable for the CEE countries is inflated. To see this, we first coded the ordered-dependent variable, *EU support*, for the same set of 13 CEE countries that Elgün and Tillman examine by using responses to the survey question posited earlier from the 2002.2 Candidate Countries Eurobarometer survey.[12] This generates the 1–3 ordered *EU support*-dependent variable described earlier.

A close examination of the ordered *EU support* variable indicates that 39% of all respondents to the survey question mentioned above opted for the middle category "neither good nor bad" response, which is indeed high. There is also good reason to suspect that this *inflated* middle category of *EU support* contains responses from two distinct sources: informed respondents *and* from a large proportion of uninformed "face-saving" respondents. To see why, first note that, prior to asking respondents about their attitude toward EU membership, the Eurobarometer 2002.2 survey evaluates whether or not respondents have any knowledge about their home country's bid for EU membership. The 2002.2 Eurobarometer survey also includes a series of nine true–false questions that objectively test respondents' knowledge about the EU as well as issues related to the impact of EU membership. Existing studies suggest that EU knowledge questions such as those described above broadly evaluate the extent of the knowledge that respondents have about the EU and the impact of EU membership (Elgün and Tillman 2007). Thus, respondents who provide several incorrect answers to the true–false questions and have not heard about their country's bid for EU membership are less likely to have sufficient knowledge about the consequences of EU membership and the EU in general.

Interestingly, we found that (1) over 75% of respondents who indicated that they had not heard of their country's bid for EU membership chose to provide an opinion on *EU support* rather than

[11]Note that the results discussed below remain unchanged when we alternatively drop all "do not know" responses.
[12]The countries included are Bulgaria, Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Romania, Slovakia, Slovenia, and Turkey (European Commission 2002b). A total of 12,147 total respondents exist in our sample.

answering "do not know," and (2) over 50% of these EU-bid uninformed respondents fall within the middle category of our *EU support* variable. Further, over 50% of respondents who *failed* to answer a *single* true–false question correctly also opted for the middle category response in *EU support*. The percentages reported above suggest that a large proportion of uninformed respondents—who lacked knowledge about the impact of EU membership and were unaware of their country's bid for EU membership—nevertheless provided the "informed" opinion that their country's (potential) membership in the EU would be "neither good nor bad." Apart from uninformed respondents, we find that a relatively smaller share of informed respondents who are cognizant about the effect of European integration and their country's bid for EU membership selected the middle category response in *EU support*. They do so based on their concrete knowledge about the EU.

The fact that the middle category of the ordered *EU support*-dependent variable is inflated and thus contains observations from two distinct sources—informed and uniformed respondents—suggests, according to our Monte Carlo analysis, that the regular OP (or ordered logit) model may not be an appropriate statistical tool for testing extant theoretical claims about public attitudes toward EU membership for especially the CEE countries. We thus turn to estimate the OP, MiOP, and MiOPC models by using the Eurobarometer survey data from 13 CEE countries—and a set of model specifications that are similar to those presented in Elgün and Tillman (2007)—to compare and contrast the estimated coefficients across these statistical models.[13]

3.2 *Covariates in the MiOP(C) Model*

The dependent variable in the OP model and in the outcome equation of the MiOP and MiOPC model is the ordered *EU support* measure which, as described above, is compiled by using responses from the 2002.2 Eurobarometer survey for 13 CEE countries. Following the theoretical literature on the determinants of *EU support* summarized above (e.g., Elgün and Tillman 2007), we first include the following three independent variables in the OP model and in the outcome equation of the MiOP and MiOPC models: *discuss politics*, *income*, and *female*.

Scholars predict and find that *discuss politics* and *income* have positive effects on *EU support*, whereas they find that the effect of *female* on *EU support* is negative, but not always statistically significant (Gabel 1998; Nelsen and Guth 2000; Carey 2002; Elgün and Tillman 2007). Hence, we anticipate that the estimates of *discuss politics* and *income* will be positive, whereas *female* will have a negative impact on *EU support*. Following Elgün and Tillman (2007), we control for a common set of variables in the OP model and in the MiOP(C) outcome stages. These controls include dichotomous variables for *unemployed*, education (*educ high*, *educ high-mid*, *educ low-mid*), *student*, occupational status (*professional*, *executive*, *manual*, *farmer*), an ordinal variable for respondents' living location (*rural*), continuous indices (0–1) for *political trust* and *xenophobia*, and a continuous measure of *age*.[14]

In the inflation equation of the MiOP and MiOPC models, we need to include covariates that predict when respondents are more (or less) likely to be informed (or knowledgable) about the EU and its impact, which may influence their attitude toward EU membership. We use insights from existing studies to identify a set of plausible covariates that can be included in the inflation equation. As this is a methods exercise, we limit our specification effort in this case by including several main variables in the inflation equation of the MiOP and MiOPC models. First, respondents may be more informed about the effect of EU membership if they communicate with their peers about European integration more frequently (Gabel 1998; Carey 2002). We therefore include *discuss politics* in the inflation equation of the MiOP and MiOPC models. Second, we add the *education* dummies to the inflation equation, as it is plausible that better-educated respondents will be more knowledgable about the EU. Third, individuals in CEE candidate countries who are aware of their country's bid for EU membership are also more likely to be informed about the EU and the

[13]The OP, MiOP, and MiOPC models have been estimated by using optim() in *R*.
[14]Operationalizations of these control variables are described in the article's Supplementary Appendix.

consequences of EU membership. We thus add the dichotomous variable *EU-bid knowledge* to the inflation equation.[15]

We next include in our inflation equations a variable for *media*, which measures how often a respondent reports watching the news on an ordinal scale. This variable may help capture whether individuals are EU-informed while also maintaining a degree of exogeneity with our main *EU support* question, given that media support for accession among CEE-candidate countries was decidedly mixed during our period of analysis (Pridham 2000, 65). We also add to our inflation stage the count variable *true EU knowledge*, which represents the number of true–false questions correctly answered by individuals on the nine-question EU-knowledge quiz discussed earlier. Since individuals who answer more questions correctly on this EU-knowledge quiz are likely to be better informed about the EU, the estimate of *true EU knowledge* should be positive. Additionally, we control for the following demographic characteristics in our inflation equations: *female*, *rural*, *student*, and *age*.

Although some of the variables listed above are incorporated in both the inflation and outcome equation (e.g., *discuss politics*), the other remaining key variables in the inflation equation, such as *media*, are—following extant research on public opinion about the EU—*not* included in the outcome equation of the MiOP(C) models.[16] This helps us address the issue of exclusion restrictions in the survey data application of the MiOP(C) models as well as adequately identify these models. Yet we also conduct a battery of specification robustness tests for these main model-specifications by, for example, adding additional theoretically relevant covariates to our inflation and outcome equations. These robustness tests are conducted to ensure that our estimated results are not driven (1) by the specific variables that we chose to include in the outcome and inflation equations and (2) by MiOP(C) under identification. The results from these robustness tests are discussed below.

3.3 *The Results*

The estimates for our OP, MiOP, and MiOPC models of *EU support* appear in Table 1.[17] We begin our assessment of these three models by first employing a number of relevant model selection statistics. To start with, we follow Harris and Zhao (2007, 1079) and employ a *t*-test of $\rho = 0$ to compare our MiOP and MiOPC models. The result from this *t*-test for our data application reveals that ρ is negative and significant, which thereby favors the MiOPC model over the MiOP model. This finding is further corroborated by likelihood ratio tests and the Akaike information criterion (AIC), which again each favor the MiOPC to the MiOP. Because the OP model is not nested within the MiOP and MiOPC models via parameter restrictions, we compare the OP model to the MiOP and MiOPC models using AIC statistics and the Vuong test for nonnested models (Vuong 1989). In both cases, these statistics favor the MiOP and MiOPC models over the OP model. Thus, the model fit statistics not only suggest that the MiOP(C) models are more appropriate for our data than the OP model, but also unambiguously favor the MiOPC over the MiOP model. We therefore primarily focus on comparing the estimates from the MiOPC and OP models below to conserve space.

At this stage, we turn to report the inflation stage results in Table 1. The estimate of *discuss politics* is positive and significant ($p < .01$) in the inflation stage of the MiOP and MiOPC models, which suggests that those who discuss politics more often are more likely to be informed about the general content and consequences of the EU membership. Our two direct measures of *EU-bid knowledge* and *true EU knowledge* are both positive and significant at the $p < .01$ level in the inflation stage of the MiOP and MiOPC models. Hence, respondents who are aware of their country's

[15]This variable is coded 1 for individuals who responded "yes" to the question. "Have you ever heard of your country's bid to become a member of the European Union?" and zero otherwise.
[16]Conversely, some covariates in the outcome equation, such as *xenophobia*, are not included in the inflation equation.
[17]See Bagozzi and Mukherjee (2012) for replication materials.

Table 1 Main OP, MiOP, and MiOPC models of EU membership support among candidate countries (2002)

	OP	*s.e.*	*MiOP*	*s.e.*	*MiOPC*	*s.e.*
Outcome stage:						
Political trust	0.759****	(0.040)	0.904****	(0.052)	0.848****	(0.049)
Xenophobia	−0.496****	(0.046)	−0.576****	(0.053)	−0.528****	(0.051)
Discuss politics	0.077****	(0.019)	0.019	(0.024)	−0.030	(0.027)
Professional	−0.047	(0.066)	−0.087	(0.076)	−0.088	(0.073)
Executive	0.125	(0.087)	0.120	(0.101)	0.116	(0.099)
Manual	−0.120****	(0.041)	−0.129****	(0.047)	−0.123****	(0.045)
Farmer	−0.028	(0.081)	−0.051	(0.091)	−0.042	(0.087)
Unemployed	0.090**	(0.047)	0.118***	(0.056)	0.109	(0.053)
Rural	−0.014	(0.016)	0.009	(0.020)	0.028	(0.022)
Female	−0.072****	(0.026)	0.027	(0.033)	0.092***	(0.037)
Age	−0.002***	(0.001)	−0.002	(0.001)	−0.001	(0.001)
Student	0.093	(0.065)	0.146**	(0.082)	0.166**	(0.088)
Income	0.066****	(0.006)	0.073****	(0.007)	0.068****	(0.006)
Educ high	0.088**	(0.050)	0.093	(0.061)	0.102	(0.066)
Educ high-mid	0.054	(0.055)	0.009	(0.069)	0.059	(0.075)
Educ low-mid	−0.086***	(0.038)	−0.034	(0.047)	0.028	(0.051)
Inflation stage:						
Constant	–		0.426***	(0.216)	0.576****	(0.199)
Discuss politics	–		0.209****	(0.048)	0.187****	(0.043)
Rural	–		−0.088***	(0.039)	−0.082***	(0.035)
Female	–		−0.392****	(0.082)	−0.330****	(0.069)
Age	–		−0.006****	(0.002)	−0.006****	(0.002)
Student	–		−0.363***	(0.151)	−0.301***	(0.144)
EU-bid knowledge	–		0.493****	(0.102)	0.395****	(0.089)
True EU knowledge	–		0.148****	(0.021)	0.125****	(0.018)
Media	–		0.056***	(0.028)	0.044**	(0.023)
Educ high	–		−0.220	(0.136)	−0.198**	(0.116)
Educ high-mid	–		−0.515****	(0.135)	−0.446****	(0.121)
Educ low-mid	–		−0.478****	(0.092)	−0.431****	(0.082)
μ_1	−0.779****	(0.090)	−0.549****	(0.112)	−0.613****	(0.114)
μ_2	0.375****	(0.090)	0.262***	(0.110)	0.138	(0.121)
ρ	–		–		−0.747****	(0.129)
No. Obs.	9,113		9,113		9,113	

Note. *** indicates $p<.01$; ** indicates $p<.05$; * indicates $p<.10$.

bid for EU membership as well as respondents who have sufficiently accurate knowledge about the EU are more likely to be cognizant about the EU and the consequences of EU membership. *Media* is also positive and significant in the MiOP(C) models, indicating that individuals who pay more attention to the news on TV are more likely to be EU-informed. However, we find mixed, and often negative and significant, results for our education dummies. Although they are somewhat fragile to specification (see the Supplementary Appendix), these negative education-results may corroborate the findings of Sturgis, Roberts, and Smith (2010), which suggest that—*conditional on an individual being uninformed*—increasing education levels raises the pressure felt by respondents to save face. Although we merely control for the *female* dummy variable, we find that the estimate of *female* is negative and significant at the $p<.01$ level in the inflation equation. This is consistent with (1) extant findings that report that women are generally less likely to be informed about the EU (Nelsen and Guth 2000; European Commission 2002a) and (2) political-survey research on gender-based response set effects (Mondak and Anderson 2004). The estimates of *rural*, *age*, and

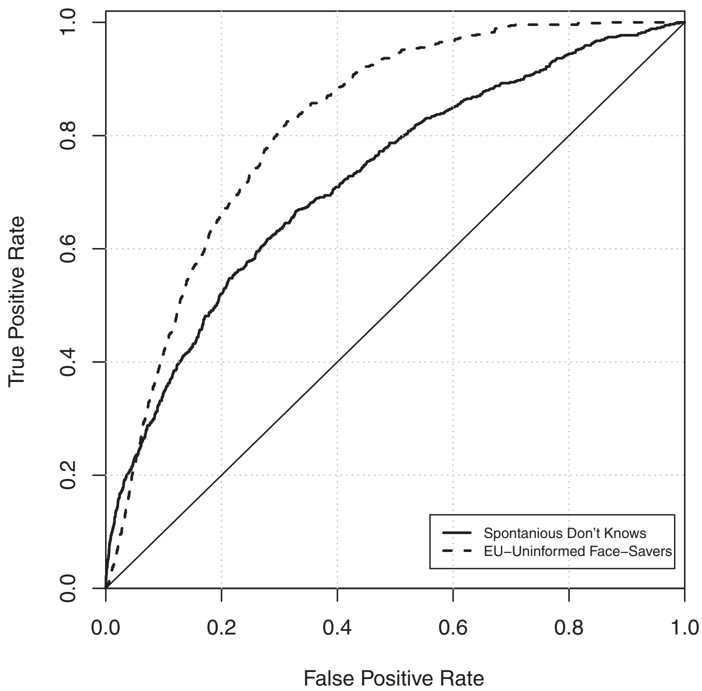

Fig. 1 ROC curves of MiOPC *pr(Inflated)* and uninformed respondents.

student are negative and significant in the inflation stage of the MiOPC models, in line with our expectations.

We can also use the inflation equation results to extract substantively rich information about (1) the probability with which uninformed respondents opt for the middle category of "neither good nor bad" in *EU support* to save face and (2) the proportion of uninformed respondents who selected the middle category. We derive the information mentioned above in three steps. First, given that the model fit statistics favor the MiOPC over the MiOP model for our particular application, we use our MiOPC inflation stage estimates to calculate the in-sample predicted probabilities of an observation being inflated. Second, we construct two binary indicators of "inflation responses" for use in evaluating the predictive accuracy of the in-sample predicted probabilities described in step one.

For the first indicator, we code a binary variable set equal to one for middle category responses that were actually "spontaneous do not knows," but were added to the middle category in accordance with past research practices (e.g., Elgün and Tillman 2007). It is coded zero otherwise. For the second, we use two additional Eurobarometer 2002.2 survey questions to construct a latent, binary indicator of *EU-uninformed face-savers*. Specifically, *EU-uninformed face-savers* is set equal to one for individuals who *both* (1) subjectively report their EU-knowledge as being greater or equal to three on a (1–10) EU-knowledge scale *and* (2) fail to answer more than two true–false questions correctly on an objective nine-question battery of EU-knowledge questions; and zero otherwise. Third, we then use *pr(inflation)* to calculate receiver operating characteristic curves (ROC) and the corresponding areas under the curve (AUC) for *spontaneous do not knows* and *EU-uninformed face-savers*, treating these latter two variables as our dependent variables and *pr(inflation)* as our predictions thereof.

The two ROC curves described above appear in Fig. 1. Note that for both latent-dependent variables, our ROC curves are well above the 50% lower bound, and return AUCs of 72% and 82% for *spontaneous do not knows* and *EU-uninformed face-savers*, respectively. Further, given that *spontaneous do not knows* and *EU-uninformed face-savers* are both (1) not actually included on the left-hand side of our inflation equation (and thus not constrained in AUC by the usual lower bound of 50%, but rather by a bound of 0%) and (2) only proxies of our intended construct, it is impressive that these AUCs are well above the traditional forecasting cutoffs for "good" predictive model accuracy. Hence, what we learn from the ROC curves and AUC results is that the inflation stage of our MiOPC model is doing a fairly accurate job of predicting the likelihood that the middle category of *EU Support* is inflated because of uninformed respondents who choose this category to save face. Moreover, the percentage effects reported above indicate quite *precisely* that a substantially high proportion of uniformed face-saving respondents opt for the middle category response of "neither good nor bad." This further validates the use of MiOPC model for our analysis, as the inflated middle category of *EU support* undoubtedly contains uninformed "face-saving" responses.

We turn to comparing the OP, MiOP, and MiOPC (ordered) outcome-stage results in Table 1. We focus on discussing the results for the following covariates that are employed to test extant theories (summarized earlier) about attitudes toward EU membership: *discuss politics*, *income*, and *female*. As predicted by existing studies,[18] the effect of both *discuss politics* and *income* on *EU support* are positive and significant in the OP model. However, while *income* remains positive and significant in the MiOPC model, the impact of *discuss politics* on *EU support* is now negative and statistically insignificant in the outcome equation of the MiOPC model. This suggests that the predicted positive effect of *discuss politics* becomes insignificant once the middle category inflation is accounted for—and the same can be said for *unemployed*. Likewise, consider the estimate of *female*. The effect of *female* on *EU support* is negative and significant in the OP model, which corroborates extant claims that women tend to have a negative attitude toward the EU (Nelsen and Guth 2000; Carey 2002). However, the effect of *female* on *EU support* is positive and significant in the outcome equation of the MiOP(C) models. Therefore—and in contrast to extant findings—females appear *more* likely to directly support EU-integration than males, once middle category inflation is accounted for.

The differences in the results obtained across the OP, MiOP, and MiOPC models for the three key covariates described above are indeed intriguing. However, to gain a better sense of the *substantive differences* that we find between our OP and MiOP(C) model estimates, we next calculate and present a number of marginal effects. Specifically, for a given variable, we compare (1) that variable's OP-predicted marginal effects to (2) that variable's MiOPC-predicted marginal effects when the variable is increased only in the outcome stage of the MiOPC model but held constant in the inflation stage.[19] These comparisons elucidate the differences between the OP marginal effects (which always return an aggregate affect) and the MiOPC direct effects (now partitioned from any inflation processes within one's aggregate sample) on *EU support*. Because of space constraints, we primarily focus on reporting the marginal effect of two variables, *discuss politics* and *female*, on the probability of observing each of the three ordered outcomes of interest for *EU support*.

The distributions of the first set of marginal effects—reporting the effect of a 1-to-3 change in *discuss politics* for each outcome of *EU support*—are presented in Fig. 2. Figure 2 indicates that the observed positive significant relationship between *discuss politics* and *EU support* in the OP model is being driven, in large part, by the positive (inflation) effects of *discuss politics* on an individual being more informed about EU integration. However, when the latter phenomenon is held constant, as in the case of the MiOPC distribution in Fig. 2, we instead observe that there is not a positive significant direct effect of *discuss politics* on support for EU integration, but rather an indeterminant direct effect. This finding suggests that individuals who discuss politics more frequently are no more or less likely to be pro-EU, but rather are simply more likely to provide an informed response.

The second set of marginal effects, reporting the effect of a 0-to-1 change in *female*, are presented in Fig. 3, and suggest that in the aggregate, *female* has a significant negative relationship with *EU support* in the OP specification, which is consistent with existing theoretical claims (Carey 2002). Indeed, these marginal effects predict that moving from male to female will make an individual 4% less likely to respond that EU membership would be "a good thing." However, as Fig. 3 indicates,

[18]See, e.g., Gabel (1998); Carey (2002); Elgün and Tillman (2007).

[19]All other variables are held to their means or modes, marginal effects were calculated using parametric bootstraps with $m = 1.000$.

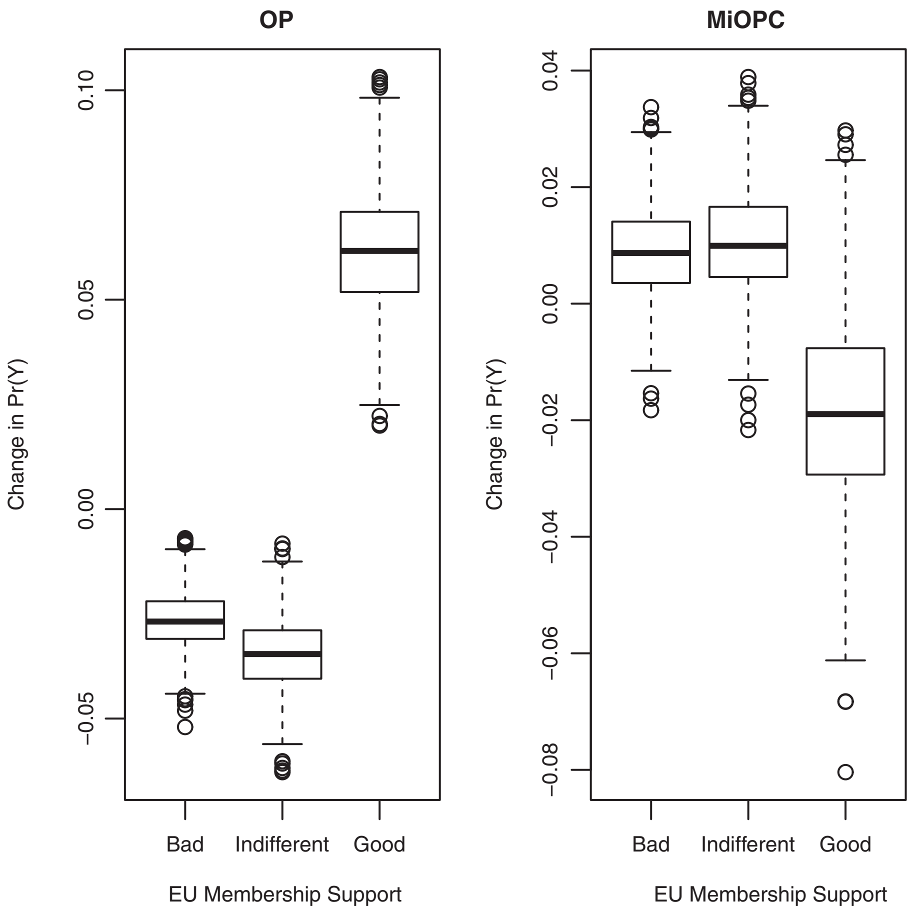

Fig. 2 Change in predicted *pr*(*EU Support*), given a 1-to-3 change in *Discuss Politics*.

this negative relationship is being driven almost entirely by the negative effect that *female* has on an individual's likelihood of being (un)informed about the EU. When this inflation effect is held constant, as in the MiOPC results reported in Fig. 3, we see that *females* who are informed about the EU are actually an average of 4% *more* likely to be positively predisposed toward the EU, relative to males. Finally, note that the marginal effects of *discuss politics* and *female* that are reported in Table A.9 of the Supplementary Appendix reinforce the substantive effects discussed here, as in each case there is a sizable and significant difference in the marginal effects of these two variables within our OP and MiOPC models.

Therefore, we can conclude that a number of commonly studied predictors of *EU support* become insignificant, or completely reverse in sign, once one accounts for middle category inflation in respondents' answers. The results obtained for the remaining controls, however, largely support existing claims about their effect(s) on *EU support*. For example, the effect of *xenophobia* is in the predicted negative direction and is highly significant, while *political trust* and *income* are in the predicted positive direction and are significant in the OP, MiOP, and MiOPC models, all of which are in line with the findings reported by Elgün and Tillman (2007).

Finally, we have conducted a battery of specification robustness tests to verify that the key findings discussed above are robust to variable choices and model-identification issues. For instance, it is plausible that *income* may have a statistically important effect in the inflation equation, while the *true-EU knowledge* measure described earlier could have a positive effect on *EU support* in the outcome equation. We thus added *income* to the inflation equation and *true-EU knowledge* to the outcome equation of the MiOP(C) models. As shown in Table A.6 in the Supplementary Appendix, the results reported above remain robust in the MiOP(C) models after adding the two covariates along the lines mentioned above. We also checked whether our results

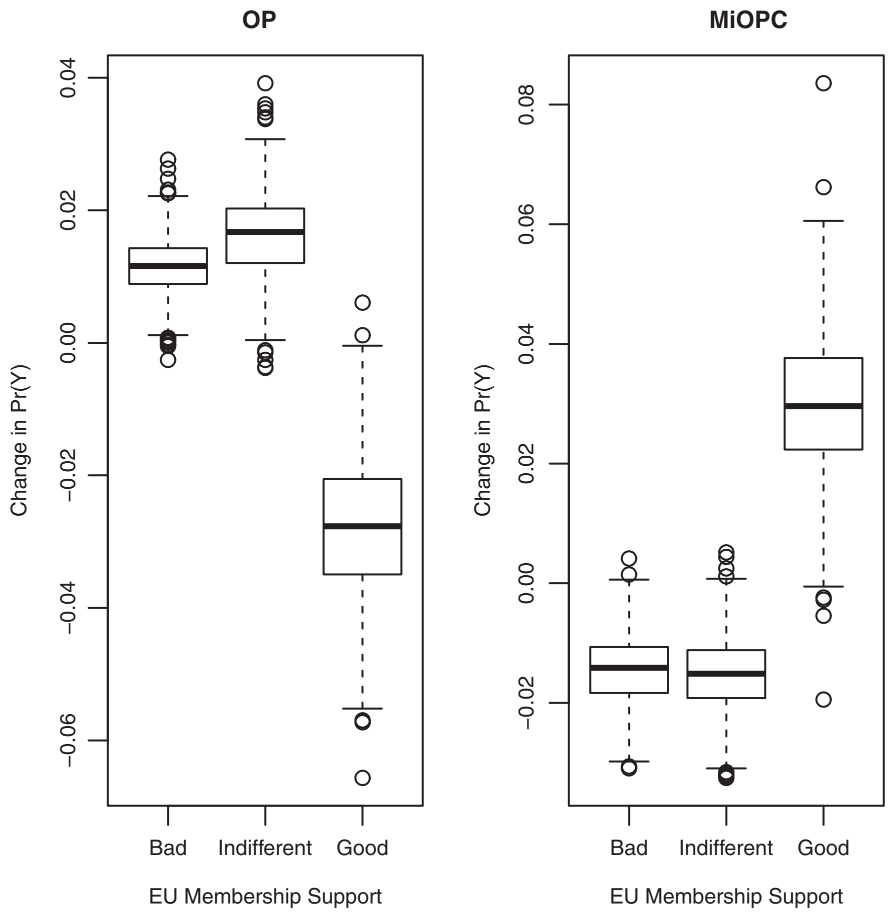

Fig. 3 Change in predicted *pr*(*EU Support*), given a 0-to-1 change in *Female*.

hold when we replaced in the inflation equation the (1) *media* attention measure given by total-TV-news attention with a direct measure of political-economic news-media attention and (2) *EU-bid knowledge* with a measure of EU-accession knowledge. Finally, we have also evaluated the robustness of our results within a series of baseline and intermediate specifications where we include all of the independent variables and some (but not all) of the control variables discussed earlier. Our results herein are highly consistent across all alternative specifications mentioned above (see Tables A.3–A.8 in the Supplementary Appendix). This suggests that the results from the MiOP(C) models are robust, and for this particular application, likely identified.[20]

7 Conclusion

The middle category of ordered survey–response-dependent variables may in some cases be "inflated" and thus generated from two distinct sources: informed respondents and uninformed (face-saving) respondents. To address the issue of "middle inflation" in ordinal-dependent variables, this article contributes to extant research on mixture models[21] by presenting the MiOP model with and without correlated errors. Results from extensive MC experiments *and* an empirical

[20]We estimated the MiOP(C) models on the following two additional survey data applications that are described in the Supplementary Appendix: individuals' support for new social movements and individuals' support for free-trade. As shown in the Supplementary Appendix, the estimates from these two additional applications (see Supplementary Tables A.10 and A.11) further demonstrate the robustness of our MiOP and MiOPC models.

[21]Jackson (1993); Harris and Zhao (2007); Xiang (2010); Imai and Tingley (2012).

application reveal that—when an ordinal-dependent variable is middle inflated—the MiOP and MiOPC models yield estimates that are superior in coverage probabilities and accuracy to OP-model estimates. Our MiOP(C) estimators also provide researchers with an opportunity to identify and include variables in not merely the outcome OP equation of the model but also the inflation equation. Doing so permits scholars to statistically account for respondents' likelihoods (1) of having (or not having) sufficient knowledge about the issue in question and (2) of subsequently opting for either the middle category or the remaining categories in the ordered survey–response-dependent variable. As a result of this nuanced framework, the marginal effects that are derived from our MiOP(C) models often reveal substantively rich insights about the impact of the key covariates on one's outcome probabilities. Finally, the estimates from the split probit equation of the MiOP(C) model yield precise information about the proportion of uninformed respondents who may choose a middle category response as a face-saving tactic. This could arguably provide substantively interesting insights that may have implications for theoretical research.

This study can be extended in four main directions. First, the statistical framework presented here can be used as a foundation to develop an inflated multinomial logit model. Doing so may be particularly useful for American politics scholars interested in assessing the determinants of vote-choice within heterogeneous populations of potential voters and never-voters.[22] American politics surveys often ask individuals to indicate which specific candidate they voted for (e.g., "Republican, Democrat, or abstained?"), and scholars then analyze these responses using multinomial logit (MNL) models of vote-choice (e.g., Arceneaux and Kolodny 2009). Misreporting is widespread in these contexts, and methods have recently been developed to address this problem in the binary setting (Katz and Katz 2010). However, even when reporting is perfect, "vote-abstention" responses likely arise from two distinct sources. Some nonvoters are likely to be completely disengaged "never-voters," who pay no regard to particular elections or candidates in their decisions to routinely abstain from voting. Many other abstainers, however, will be "potential-voters" who nevertheless abstained from a given election cycle due to temporary political or economic factors such as a distaste for *all* of the candidates running in a particular election. Including both sets of nonvoters within an MNL model of vote-choice could lead one to underestimate the direct effects of treatments or exogenous shocks on candidate-selection or turnout. An inflated MNL model would account for this particular form of "inflation" in a nuanced and unbiased manner.

Second, apart from misreporting by "uninformed" individuals in survey questions of the attitude or opinion variety, respondents also tend to report that they have voted even when they do not do so (Katz and Katz 2010). The methodological problems that emerge from such misreporting have been thoroughly addressed by Katz and Katz (2010). As a supplement to Katz and Katz's (2010) findings, we suggest—based on this study—that it may be worthwhile to develop a two-stage-inflated probit model that accounts for the likelihood with which voters misreport their turnout decision and the factors that determine their turnout decision. Third, a useful direction for future research will involve application of the MiOP(C) models to survey studies of immigration-support, future economic-assessment, and political ideology, as these studies have also reported high proportions of "indifference" (i.e., middle category) responses. Applying the MiOP(C) model to these survey studies may yield interesting empirical insights since it is plausible that a high share of uninformed individuals may have responded to survey questions about immigration-support, future economic assessment, and ideology by opting for the middle category of indifference. Fourth, studies on survey respondents in the United States have shown that respondents who are more uncertain about their attitudes on issues like abortion or tax policy are more likely to place their response at the middle category of the ordered scale on questions that track their perceptions about their senator's (1) ideological position and (2) position on abortion and tax policy (Alvarez and Franklin 1994a, 1994b). Building on these studies, the application presented above provides valuable insights into how uncertainty influences middle category response-set behaviors within a similar political-knowledge area: support for EU-membership. Hence, the

[22]Or, for International Relations scholars interested in war-joining or exchange-rate regime choices.

MiOPC model can be potentially extended to jointly assess when respondents are more likely to be uncertain about certain policy issues and how this in turn affects their perceptions about the incumbent senator's position on these issues.

References

Alvarez, R. M., and C. H. Franklin. 1994a. Attitudes, uncertainty and survey responses. Working Paper, Cal Tech.

———. 1994b. Uncertainties and political perceptions. *Journal of Politics* 56(3):678–88.

Arceneaux, K., and R. Kolodny. 2009. Educating the least informed: Group endorsements in a grassroots campaign. *American Journal of Political Science* 53(4):755–70.

Bagozzi, B. E., and B. Mukherjee. 2012. Replication data for: A mixture model for middle-category inflation in ordered survey responses. IQSS Dataverse Network: http://hdl.handle.net/1902.1/18183 (last accessed 20 June 2012).

Barabas, J. 2002. Another look at the measurement of political knowledge. *Political Analysis* 10(2):1–14.

Berinsky, A. J. 2002. Political context and the survey response: The dynamics of racial policy opinion. *Journal of Politics* 64(2):567–84.

Bound, J., C. Brown, and N. Mathiowetz 2001. Measurement error in survey data. In *Handbook of Econometrics*, eds. J. J. Heckman and E. E. Leamer, 3705–843. Oxford, UK: Elsevier.

Brady, H. E., and P. M. Sniderman 1991. The likability heuristic. In *Reasoning and choice: Explorations in political psychology*, eds. P. M. Sniderman, R. A. Brody, and P. E. Tetlock. Cambridge, UK: Cambridge University Press.

Burden, B. C. 2000. Voter turnout and the national election studies. *Political Analysis* 8(4):389–98.

Burger, M., F. van Oort, and G.-J. Linders. 2009. On the specification of the gravity model of trade: Zeros, excess zeros, and zero-inflated estimation. *Spatial Economic Analysis* 4(2):167–90.

Carey, S. 2002. Undivided loyalties: Is national identity an obstacle to European integration? *European Union Politics* 3(4):387–413.

Christin, T. 2005. Economic and political basis of attitudes toward the EU in Central and East European countries in the 1990s. *European Union Politics* 6(1):29–57.

Delli Carpini, M. X., and S. Keeter. 1997. *What do Americans know about politics and why it matters*. New Haven, CT: Yale University Press.

Elgün, O., and E. R. Tillman. 2007. Exposure to European Union policies and support for membership in the candidate countries. *Political Research Quarterly* 60(3):391–400.

European Commission. 2002a, Spring. Eurobarameter 57: Eu 15 report. 50.

———. 2002b, September–October 2002. Candidate countries Eurobarometer 2002.22.

Ferber, R. 1956. The effect of respondent ignorance on survey results. *Journal of the American Statistical Association* 51(276):576–86.

Fiske, S. T., R. R. Lau, and R. A. Smith. 1990. On the varieties and utilities of political expertise. *Social Cognition* 8:31–48.

Gabel, M. 1998. Public support for European integration: An empirical test of five theories. *Journal of Politics* 60(2):333–54.

Gerber, A. S., and G. A. Huber. 2010. Partisanship, political control, and economic assessments. *American Journal of Political Science* 54(1):153–73.

Gerber, A. S., G. A. Huber, D. Doherty, C. M. Dowling, and S. E. Ha. 2010. Personality and political attitudes: Relationships across issue domains and political contexts. *American Political Science Review* 104(1):111–33.

Hainmueller, J., and M. J. Hiscox. 2010. Attitudes toward highly skilled and low-skilled immigration: Evidence from a survey experiment. *American Political Science Review* 104(1):61–84.

Harris, M. N., and X. Zhao. 2007. A zero-inflated ordered probit model, with an application to modeling tobacco consumption. *Journal of Econometrics* 141(2):1073–99.

Heckman, J. J. 1979. Sample selection bias as a specification error. *Econometrica* 47(1):153–61.

Heckman, J. J., and G. L. Sedlacek. 1990. Self-selection and the distribution of hourly wages. *Journal of Labor Economics* 8(1):329–63.

Hooghe, L., and G. Marks. 2005. Calculation, community, and cues: Public opinion on European integration. *European Union Politics* 6(4):419–43.

Imai, K., and D. Tingley. 2012. A statistical method for empirical testing of competing theories. *American Journal of Political Science* 56(1):218–36.

Inglehart, R. 1970. Cognitive mobilization and European identity. *Comparative Politics* 3(1):45–70.

Jackson, J. E. 1993. Attitudes, no opinions, and guesses. *Political Analysis* 5(1):39–60.

Karp, J. A., S. A. Banducci, and S. Bowler. 2003. To know it is to love it? Satisfaction with democracy in the European Union. *Comparative Political Studies* 36(3):271–91.

Katz, J. N., and G. Katz. 2010. Correcting for survey misreports using auxiliary information with an application to estimating turnout. *American Journal of Political Science* 54(3):175–93.

Kuklinski, J. H., P. J. Quirk, J. Jerit, D. Schwieder, and R. F. Rich. 2000. Misinformation and the currency of citizenship. *Journal of Politics* 62(3):790–816.

Mondak, J. J. 2000. Reconsidering the measurement of political knowledge. *Political Analysis* 8(1):57–82.

Mondak, J. J., and M. R. Anderson. 2004. The knowledge gap: A reexamination of gender-based differences in political knowledge. *Journal of Politics* 66(2):492–512.

Mondak, J. J., and B. Creel Davis. 2001. Asked and answered: Knowledge levels when we won't take "don't know" for an answer. *Political Behavior* 23(1):199–224.

Nelsen, B. F., and J. L. Guth. 2000. Exploring the gender gap: Women, men, and public attitudes toward European integration. *European Union Politics* 1(3):267–91.

Pridham, G. 2000. EU accession and domestic politics: Policy consensus and interactive dynamics in Central and Eastern Europe. *Journal of Development Economics* 1(1):49–74.

Richey, S. 2010. The impact of anti-assimilationist beliefs on attitudes toward immigration. *International Studies Quarterly* 54(1):197–212.

Rohrschneider, R. 1990. The roots of public opinion toward new social movements: An empirical test of competing explanations. *American Journal of Political Science* 34(1):1–30.

———. 2002. The democracy deficit and mass support for an EU-wide government. *American Journal of Political Science* 46(2):463–75.

Sturgis, P., C. Roberts, and P. Smith. 2010. Middle alternatives revisited: How the neither/nor response acts as a "face-saving" way of saying "I don't know." Working Papers, Southampton Statistical Sciences Research Institute Methodology.

Tucker, J. A., A. C. Pacek, and A. J. Berinsky. 2002. Transitional winners and losers: Attitudes toward EU membership in post-Communist countries. *American Journal of Political Science* 46(3):557–71.

Vuong, Q. H. 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57(2):307–33.

Winkelman, R. 1998. Count data models with selectivity. *Econometric Reviews* 17(4):339–59.

Xiang, J. 2010. Relevance as a latent variable in dyadic analysis of conflict. *Journal of Politics* 72(2):484–98.

Zaller, J. 1992. *The nature and origins of mass opinion*. Cambridge, MA: Cambridge University Press.