The Annals of Applied Statistics
2015, Vol. 9, No. 4, 2237–2265
DOI: 10.1214/15-AOAS878

ASSESSING NONRESPONSE BIAS IN A BUSINESS SURVEY: PROXY PATTERN-MIXTURE ANALYSIS FOR SKEWED DATA

BY REBECCA ANDRIDGE AND KATHERINE JENNY THOMPSON

Ohio State University and U.S. Census Bureau

The Service Annual Survey (SAS) is a business survey conducted annually by the U.S. Census Bureau that collects aggregate and detailed revenues and expenses data. Typical of many business surveys, the SAS population is highly positively skewed, with large companies comprising a large proportion of the published totals. When alternative data are not available, missing data are handled with ratio imputation models that assume missingness is at random. We propose a proxy pattern-mixture (PPM) model that provides a simple framework for assessing nonresponse bias with respect to different nonresponse mechanisms. PPM models were first introduced in this context by Andridge and Little [*Journal of Official Statistics* **27** (2011) 153–180], but their model assumed the characteristic of interest and the predicted proxy have a bivariate normal distribution, conditional on the missingness indicator. Although often appropriate for large demographic surveys, the normality assumption is less justifiable for the highly skewed SAS data. We propose an alternative PPM model using a bivariate gamma distribution more appropriate for the SAS data. We compare the two PPM models through application to data from six years of data collection in three industries in the health care and transportation sectors of the SAS. Finally, we illustrate properties of the method through simulation.

1. Introduction. Effects of nonresponse on bias in survey estimates have been studied extensively over the past decade, as response rates to large-scale surveys continue to decline while public and federal demands for timelier and more detailed measures continue to emerge. Nonresponse can occur for an entire survey unit (unit nonresponse) or for selected items provided by a survey unit (item nonresponse). Our research is motivated by the Service Annual Survey (SAS) conducted by the U.S. Census Bureau, a business survey that uses imputation to account for unit and item nonresponse. Typical of many business surveys, the SAS population is highly positively skewed, with large companies comprising a large proportion of the published totals. The SAS imputation procedures are designed to exploit available information when there is nonresponse in order to obtain inference about population parameters by maximizing the usage of available reliable auxiliary data for substitution. When such data are not available, then the SAS imputes replacement values via ratio imputation models that rely on strong (verifiable) linear rela-

Received March 2015; revised July 2015.
Key words and phrases. Missing data, nonresponse bias analysis, nonignorable missingness, multiple imputation, skewed data, business surveys, proxy pattern-mixture models.

tionships, especially with total revenue and total expenses, the two key published items.

The Office of Management and Budget requires that federal programs perform nonresponse bias analyses when the unit response rate falls below 80% or response rates for published items fall below 70%. Of course, the literature supports the hypothesis that nonresponse bias is not necessarily a function of survey response rates; for example, see Peytcheva and Groves (2009). Indeed, although the SAS unit response rates tend to fall below 70%, the survey methodologists have long contended that their item-specific imputation procedures mitigate the effects of nonresponse bias on the survey's key estimates. However, this contention is entirely anecdotal. The proxy pattern-mixture model analysis approach described below provides a statistical method to evaluate/assess this hypothesis separately for each survey item in a nonresponse bias analysis study.

There are three components that can be used to assess the potential for nonresponse bias: the amount of nonresponse, the differences between respondents and nonrespondents on fully observed characteristics (e.g., paradata, frame data), and the relationship between these fully observed characteristics and the survey outcomes (only measurable among respondents). Wagner (2012) describes a typology that classifies types of "indicators" of nonresponse bias based on these three components: (1) indicators based on response propensities, (2) indicators involving the response propensity and fully observed characteristics and (3) indicators involving the response propensity, fully observed characteristics and the survey outcome. Andridge and Little (2011) demonstrate a proxy pattern-mixture (PPM) analysis approach to assessing the potential for nonresponse bias that falls under the third category, assimilating all three components into a single sensitivity analysis. Their framework facilitates assessment of nonresponse bias with respect to different nonresponse mechanisms. In particular, their PPM analysis provides estimates of survey outcomes (e.g., means) under a missing at random (MAR) assumption and various missing not at random (MNAR) assumptions, providing an assessment of the sensitivity of estimates to nonignorable missingness.

In brief, the PPM model reduces a set of fully observed auxiliary variables to a single "proxy" variable X. The joint distribution of a survey outcome Y and this proxy X is modeled as a bivariate normal distribution with separate parameters for respondents and nonrespondents (a pattern-mixture model). Through assumptions on the missing data mechanism, adjusted estimates of the mean of Y under these various missingness mechanisms are obtained. For more details, see Andridge and Little (2011).

As mentioned above, the SAS survey methodologists were very interested in using the PPM approach on their data to—hopefully—validate their contention of minimal detrimental effects on key survey estimates due to nonresponse after all adjustment procedures are completed. In particular, they were particularly interested in using the resulting values of the fraction of missing information (FMI) estimated via the PPM model as the study metric, since the FMI is easily interpretable and bounded; see Section 3.3. However, the methods presented in Andridge and Little (2011) assume that the characteristic of interest Y and the proxy X have a *bivariate normal distribution*, conditional on the missingness indicator. This normal PPM model is appealing for its computational tractability and is often appropriate for large demographic survey applications. However, it is less justifiable for highly positively skewed populations, such as business populations like the SAS. The normal PPM model is relatively robust to departures from normality since it relies on first and second moments in estimating the mean of Y. However, while estimates of the mean of Y may be robust, variance estimates may not be. For example, suppose that units with larger values of X are more likely to respond than units with smaller values of X. If the proxy model holds for respondents, variance estimates under the normal model will be inflated, which in turn could lead to misleading interpretation of the effects of nonresponse bias on the survey estimates. Thus, we sought an alternative model that could be used in the PPM framework that would incorporate two key features of the SAS business survey data (and by extension, could be applicable to other, similar business surveys): (1) skewed marginal distributions, and (2) larger variance for larger values of the proxy X.

This paper presents the results of a feasibility study conducted for the SAS using a subset of surveyed industries and data items. The primary objective of the study was to develop a variation of an accepted analysis procedure that could be easily applied to skewed business survey data; if feasible, one would expect the new method to be applied on a much larger scale to other industries. In this paper, we develop a PPM model using a bivariate gamma model. Many other studies have successfully modeled the marginal distributions of business data populations as gamma [Krewski and Rao (1981); Haziza et al. 2010; Thompson (2005)], so the choice of distribution is well supported. First, to motivate the gamma PPM model, we give an overview of the SAS in Section 2. Section 3 introduces the proposed gamma PPM model, and in Section 4 we apply both the original (normal) and bivariate gamma PPM models to empirical data from six separate years of SAS data collection in three industries located in the health care and transportation sectors. The conflicting results between the normal and bivariate gamma PPM models motivate the simulation study presented in Section 5. Some concluding remarks follow in Section 6.

2. Background on the service annual survey. The SAS is a mandatory survey of approximately 72,000 employer businesses having one or more establishments located in the U.S. that provide services to individuals, businesses and governments; the survey coverage includes most personal, business, automotive, amusement and recreation, social welfare, health care and other professional services industries. The SAS collects aggregate and detailed revenues and expenses, e-commerce, exports and inventories data from a stratified sample of business firms with paid employees in selected industries. For processing purposes, the SAS is divided into five sections, each covering one or more NAICS service sectors. At the survey methodologists' suggestion, the feasibility study is restricted to the SAS sectors covering the transportation and health industries (SAS-T and SAS-H, respectively).

The SAS uses a single-stage stratified random sample design. Companies are stratified by their major kind of business (determined by the industry containing the largest portion of total receipts for the company) and then are further sub-stratified by annual receipts or revenue (the measure of size). Each company's frame measure of size is compared to an industry-specific size cutoff. Companies whose values exceed the applicable thresholds are designated as certainty units (sampled with probability equal to one). The Employer Identification Numbers (EINs) of the remaining companies are further stratified by major kind of business and sub-stratified by the frame value of their total annual receipts or revenue. Within each noncertainty stratum, a simple random sample of EINs is selected without replacement. Thus, the sampling units are either companies or EINs. Each sampling unit represents one or more establishments/locations owned or controlled by the same firm. The initial sample is updated quarterly to reflect births and deaths, adding new employer businesses identified in the Business and Professional Classification Survey and dropping firms and EINs that are no longer active. Information on the SAS design and methodology is available at http://www.census.gov/services/sas/about_the_surveys.html.

Target populations for business surveys tend to have highly positively skewed characteristics, and the SAS is no exception. Figure 1 presents a truncated histogram of frame measure of size from the "All Other Miscellaneous Ambulatory

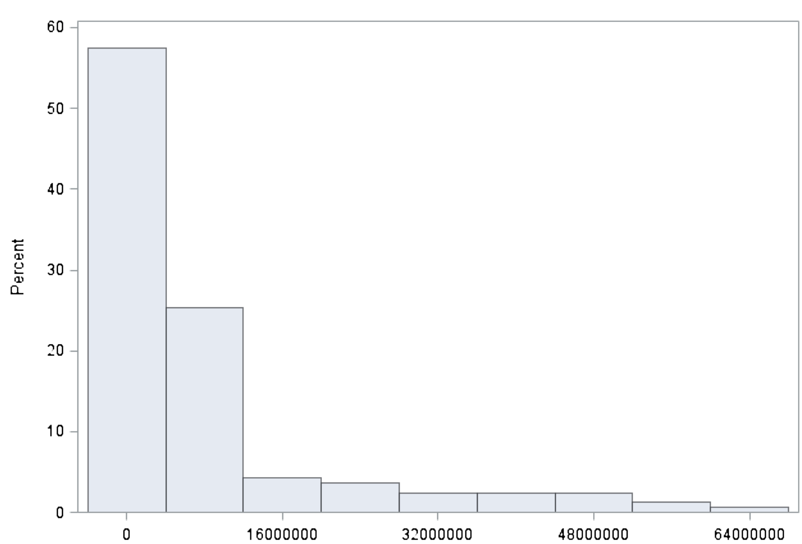

FIG. 1. *Truncated SAS-H histogram of measure of size in the "All Other Miscellaneous Ambulatory Health Care Services" Industry* (*NAICS* 621999), *with the largest cases excluded from presentation to preserve confidentiality.*

Health Care Services" industry (NAICS 621999) in the health services component of the Service Annual Survey (SAS-H). The skewness coefficient (calculated on the nontruncated data) is 2.8, indicative of positive skew. The larger businesses (those in the right-hand tail) are more likely than small business to provide response data. First, the smaller units may not keep track of the requested data items [Willimack and Nichols (2010) and Snijkers et al. (2013), Chapter 2] or may perceive the response burden as being quite high [Bavdaž (2010)]. Second, to improve or maintain the quality of the totals, operational procedures are designed to increase the likelihood of obtaining valid response from large units, for example by performing phone call follow-up of the largest cases, followed by intensive analyst research for auxiliary data sources such as publicly available financial reports to replace imputed values with equivalent data [Thompson and Oliver (2012)].

Imputation methodology is used to account for both unit and item nonresponse in the SAS. These models use auxiliary survey and administrative records data as input. The imputation cells are six-digit industry (NAICS) code cross-classified by tax-exempt status. Unlike the sampling strata definitions, the imputation cells do not account for unit size (in terms of expected value of total receipts), and imputation parameters use certainty and (weighted) noncertainty units within the same cell. The imputation base for the ratio imputation parameters is restricted to complete respondent data, subject to outlier detection and treatment.

To perform item imputation, direct substitution of administrative or auxiliary survey data is implemented as feasible; this is referred to as a logical edit procedure. Otherwise, the SAS uses ratio imputation to account for nonresponse. The ratio imputation model assumes that the finite population of y is generated from a superpopulation model m such that $E_m(y|x) = \beta x$ and $V_m(y|x) = x\sigma^2$, where x is a strictly positive continuous variable and y is drawn from an unspecified distribution. Given a sample of size n, with r respondents and $n - r$ nonrespondents, estimates of β under this model are given by $\hat{\beta} = \bar{y}_r/\bar{x}_r$ (without design weights) or by $\hat{\beta} = \sum_{i\in R} w_i y_i / \sum_{i\in R} w_i x_i$ (with design weights), where w_i is a design weight and $i \in r$ denotes summing over respondents [Lohr (2010), Chapter 4]. The ratio imputation model requires that the regression line goes through the origin, a reasonable assumption with many business data items (e.g., if the business does not have employees, then no payroll is expended). Thompson and Washington (2013) provide a more complete discussion on the challenges of developing unit nonresponse adjustment procedures for the SAS program.

These key features of data from business surveys—severe positive skew and the use of ratio imputation procedures—motivate the development of a new method for assessing the sensitivity of inferences to nonresponse. In particular, we sought to develop a framework similar to the proxy pattern-mixture model of Andridge and Little (2011) that could be applied to SAS and other business survey data with skewed characteristics.

3. A proxy pattern-mixture model for positively skewed data. In developing our method we adopt a model-based approach to survey sampling, by which we specify a "superpopulation model" and use this model as the basis for inference [Royall (1992)]; this is in contrast to the design-based (randomization-based) classical approach, where the outcome variables are assumed to be "fixed" in the population and inference is made with respect to the sample selection probabilities. As such, we assume that we have a simple random sample of size n from an infinite population. Let Y_i denote a continuous positive survey outcome and $Z_i = (Z_{i1}, Z_{i2}, \dots, Z_{iq})$ the values of q covariates for unit i in the sample. Only r of the n sampled units respond, so observed data consist of (Y_i, Z_i) for $i = 1, \dots, r$ and Z_i for $i = r + 1, \dots, n$.

As in Andridge and Little (2011), we start by reducing the dimension of the covariates Z by creating a *proxy variable* X using linear regression of respondent data. We regress Y on Z and take the proxy X to be the predicted values of Y from the regression model, which is thus available for both respondents and nonrespondents. We exclude the intercept in this model to create the proxy, paralleling the ratio estimation/ratio imputation models used for the SAS data. By definition, the outcome Y only takes positive values, and consequently the proxy X should as well. In the applications described in Section 4, this no-intercept model helps ensure that X is positive. In the SAS data, the covariates Z are positive (e.g., total sales, payroll or expenditures) and relationships between Z and Y are strong, so no negative predictions of Y are made. If the linear regression results in a negative proxy value for a specific data set, then our proxy pattern-mixture model is not an appropriate choice. Note that if the outcome variable can be real valued (e.g, profit/loss, income), then our gamma proxy pattern-mixture model is likewise not appropriate; however, the normal PPM of Andridge and Little (2011) may be.

We note that while the construction of the proxy uses regression parameters estimated using respondent data only, in subsequent estimation and inference we do not have to assume that the parameters of the regression of Y on Z are the same for both respondents and nonrespondents. Under an MAR mechanism, by definition these parameters are the same, but under MNAR mechanisms they will be different. The proxy X serves merely as a method of reducing multivariate Z to a univariate X as was done by Andridge and Little (2011). If the correlation between Y and X for respondents is high we call X a "strong proxy" for Y and if this correlation is low we call X a "weak proxy" for Y. The type of auxiliary variables Z available will affect the strength of the proxy. In household surveys, design variables and paradata have been shown to have weak correlations with survey outcomes [Kreuter et al. (2010)]. However, since our application is restricted to the key items collected by the SAS, we have strong predictors of Y [Thompson and Washington (2013)].

Our goal is to assess the potential for nonresponse bias for the SAS under varying assumptions on the missing data mechanism, specifically, when data are MAR versus MNAR. We focus on estimation of the mean of Y, but other estimators such as totals could easily be estimated under our framework. Let M denote the missingness indicator, such that $M = 0$ if Y is observed and $M = 1$ if Y is missing. The PPM model as described by Andridge and Little (2011) assumes bivariate normality for the outcome Y and proxy X conditional on M. As mentioned in Section 2, for the highly skewed characteristics of business populations such as the SAS, this assumption seems questionable. Moreover, the ratio imputation model commonly used for business data assumes that the variance of Y given X depends on X, which is not true for the bivariate normal PPM model. Instead, we assume that the joint distribution of (Y, X, M) follows a pattern-mixture model using Kibble's (1941) bivariate gamma distribution (KBGD), where respondents ($M = 0$) and nonrespondents ($M = 1$) have distinct parameters:

$$(1)\qquad \begin{aligned} &(Y, X|M = m) \sim \mathrm{KBGD}(\alpha^{(m)}, \nu_y^{(m)}, \nu_x^{(m)}, \rho^{(m)}), \\ &M \sim \mathrm{Bernoulli}(1 - \pi). \end{aligned}$$

The joint density of the $\mathrm{KBGD}(\alpha^{(m)}, \nu_x^{(m)}, \nu_y^{(m)}, \rho^{(m)})$ is given by

$$(2)\qquad \begin{aligned} &f(x^{(m)}, y^{(m)}|m, \alpha^{(m)}, \nu_x^{(m)}, \nu_y^{(m)}, \rho^{(m)}) \\ &\quad = \frac{(\nu_x^{(m)}\nu_y^{(m)})^{\alpha^{(m)}}}{(1-\rho^{(m)})\Gamma(\alpha^{(m)})}\left(\frac{xy}{\rho^{(m)}\nu_x^{(m)}\nu_y^{(m)}}\right)^{(\alpha^{(m)}-1)/2} \\ &\qquad \times \exp\left(-\frac{\nu_x^{(m)}x + \nu_y^{(m)}y}{1-\rho^{(m)}}\right) I_{\alpha^{(m)}-1}\left(\frac{2\sqrt{\rho^{(m)}\nu_x^{(m)}\nu_y^{(m)}xy}}{1-\rho^{(m)}}\right), \end{aligned}$$

where $x, y, \nu_x^{(m)}, \nu_y^{(m)}, \alpha^{(m)} > 0$, $0 \le \rho^{(m)} < 1$, and $I_{\alpha(m)}(\cdot)$ is the modified Bessel function of the first kind of order $\alpha^{(m)}$.

Under this pattern-mixture model, the marginal distributions of X and Y given M are $\mathrm{Gamma}(\alpha^{(m)}, \nu_x^{(m)})$ and $\mathrm{Gamma}(\alpha^{(m)}, \nu_y^{(m)})$, that is, gamma with shape parameter $\alpha^{(m)}$ and rate parameters $\nu_x^{(m)}$ and $\nu_y^{(m)}$ (i.e., $E[X] = \alpha/\nu_x$). The dependence between X and Y given M is described with $\mathrm{Corr}(X, Y|M) = \rho^{(m)}$. The shared shape parameter $\alpha^{(m)}$ is a feature of the KBGD; this assumption should be checked before applying our PPM model by comparing estimates of the shape parameter for Y and X among respondents ($m = 0$). If this assumption does not hold, then fitting the KBGD (e.g., by maximum likelihood) will result in biased parameter estimates. For the marginal gamma distribution with the smaller (larger) shape parameter, the estimates of both the shape and rate parameters will be biased upward (downward), resulting in an unbiased estimate of the mean, but an underestimate (overestimate) of the variance.

The conditional distribution of Y given X and M is a randomized gamma distribution [Feller (1966)], also called a Bessel function distribution of the first type by Yuan and Kalbfleisch (2000). The pattern-mixture model based on the KBGD has the attractive property that both the expectation and variance of Y given X and M are linear in X,

$$\begin{aligned} E[Y|X=x, M=m] &= \frac{\alpha^{(m)}(1-\rho^{(m)})}{\nu_y^{(m)}} + \frac{\rho^{(m)}\nu_x^{(m)}}{\nu_y^{(m)}}x = \beta_{y0.x}^{(m)} + \beta_{yx.x}^{(m)}x, \\ \operatorname{Var}[Y|X=x, M=m] &= \frac{\alpha^{(m)}(1-\rho^{(m)})^2}{\nu_y^{(m)2}} + \frac{2\rho^{(m)}(1-\rho^{(m)})\nu_x^{(m)}}{\nu_y^{(m)2}}x. \end{aligned} \tag{3}$$

The conditional expectation and variance of X given Y and M has a similar form, with the roles of the y and x terms (and associated parameters) reversed.

The KBGD was chosen over other bivariate distributions applicable to skewed data, such as the bivariate lognormal, because of this linearity. $E[Y|X=x]$ has a regression-like form amenable to adapting for a pattern-mixture model, and the dependence of $V[Y|X]$ on X reflects our beliefs about the business survey data. Additionally, this form is similar to the commonly used ratio imputation model presented in Section 2, where both the expectation and variance are linear in X, with the addition of an intercept in the bivariate gamma model. We call the proxy pattern-mixture model that results from using the KBGD the "gamma PPM" model to distinguish it from the "normal PPM" model of Andridge and Little (2011).

The pattern-mixture model defined by (1) is underidentified; there is no information in the data to estimate the nonrespondent parameters $\nu_y^{(1)}$ or $\rho^{(1)}$. However, as with previously described pattern-mixture models [e.g., Little (1994)], parameter restrictions induced by assumptions on the missing data mechanism allow the model to be identified. Specifically, we assume that the probability that Y is missing depends on a linear combination of the proxy X and the outcome Y,

$$P(M=1|Y,X) = f(X+\lambda Y).$$

Here λ is a sensitivity parameter that determines the missingness mechanism, and f is an unspecified function. When $\lambda = 0$, missingness depends only on X, and data are MAR. This is the case where the regression of Y on Z is assumed to be the same for respondents and nonrespondents. Non-zero values of λ correspond to MNAR mechanisms, since in this case missingness depends on the partially observed Y. When $\lambda = \infty$, missingness is a type of "extreme" MNAR depending only on the unobserved Y and not on X.

There is no information in the data with which to estimate λ, thus we perform a sensitivity analysis using $\lambda \in \{0, \infty\}$ to bound the potential for nonresponse bias. In their sensitivity analysis, Andridge and Little (2011) also consider an intermediate case of $\lambda = 1$, a "compromise" response mechanism which *equally* weights the contribution from the proxy X and the outcome Y. However, if X and Y follow the KBGD, the distribution of $X+Y$ is not any type of gamma distribution [Izawa (1965)], and thus we restrict ourselves to $\lambda = 0$ (MAR) and $\lambda = \infty$ ("extreme"

MNAR). We note that the intermediate case using $\lambda = 1$ (or any other value between 0 and ∞) would produce estimates that lie between the extremes of $\lambda = 0$ and $\lambda = \infty$, thus by omitting $\lambda = 1$ we are still bounding the potential for nonresponse bias.

If we assume that missingness depends only on X ($\lambda = 0$, MAR), then the conditional distribution of Y given X is the same for respondents ($m = 0$) and nonrespondents ($m = 1$), implying the restrictions $\beta^{(1)}_{y0.x} = \beta^{(0)}_{y0.x}$ and $\beta^{(1)}_{yx.x} = \beta^{(0)}_{yx.x}$ in (3). Setting these terms equal and solving for the unidentified parameters $\nu^{(1)}_y$ and $\rho^{(1)}$ yields

$$\begin{aligned} \nu^{(1)}_y &= \frac{\alpha^{(1)}\nu^{(1)}_x}{\alpha^{(1)}\beta^{(0)}_{yx.x} + \nu^{(1)}_x\beta^{(0)}_{y0.x}} = \frac{\alpha^{(1)}\nu^{(1)}_x\nu^{(0)}_y}{\alpha^{(1)}\rho^{(0)}\nu^{(0)}_x + \alpha^{(0)}(1-\rho^{(0)})\nu^{(1)}_x}, \\ \rho^{(1)} &= \frac{\alpha^{(1)}\beta^{(0)}_{yx.x}}{\alpha^{(1)}\beta^{(0)}_{yx.x} + \nu^{(1)}_x\beta^{(0)}_{y0.x}} = \frac{\alpha^{(1)}\rho^{(0)}\nu^{(0)}_x}{\alpha^{(1)}\rho^{(0)}\nu^{(0)}_x + \alpha^{(0)}(1-\rho^{(0)})\nu^{(1)}_x}. \end{aligned} \tag{4}$$

If we assume that missingness depends only on Y ($\lambda = \infty$, MNAR), the conditional distribution of X given Y is the same for respondents and nonrespondents [recall that the distribution of $X|Y, M$ is (3) with the x and y terms swapped]. Thus $\beta^{(1)}_{x0.y} = \beta^{(0)}_{x0.y}$ and $\beta^{(1)}_{xy.y} = \beta^{(0)}_{xy.y}$, and by setting these terms equal we can again solve for $\nu^{(1)}_y$ and $\rho^{(1)}$ to obtain

$$\begin{aligned} \nu^{(1)}_y &= \frac{\alpha^{(1)}\nu^{(1)}_x\beta^{(0)}_{xy.y}}{\alpha^{(1)} - \nu^{(1)}_x\beta^{(0)}_{x0.y}} = \frac{\alpha^{(1)}\rho^{(0)}\nu^{(1)}_x\nu^{(0)}_y}{\alpha^{(1)}\nu^{(0)}_x - \alpha^{(0)}(1-\rho^{(0)})\nu^{(1)}_x}, \\ \rho^{(1)} &= \frac{\alpha^{(1)} - \nu^{(1)}_x\beta^{(0)}_{x0.y}}{\alpha^{(1)}} = \frac{\alpha^{(1)}\nu^{(0)}_x - \alpha^{(0)}(1-\rho^{(0)})\nu^{(1)}_x}{\alpha^{(1)}\nu^{(0)}_x}. \end{aligned} \tag{5}$$

Once $\nu^{(1)}_y$ and $\rho^{(1)}$ have been identified, the marginal mean of Y can be obtained as is standard in a pattern-mixture model,

$$\mu_y = \pi\frac{\alpha^{(0)}}{\nu^{(0)}_y} + (1-\pi)\frac{\alpha^{(1)}}{\nu^{(1)}_y}. \tag{6}$$

3.1. *Maximum likelihood estimation.* Maximum likelihood estimates (ML) for the gamma PPM model are obtained in two steps. The first step is to use an iterative algorithm such as Newton–Raphson to obtain ML estimates for the identifiable parameters, $(\alpha^{(0)}, \nu^{(0)}_x, \nu^{(0)}_y, \rho^{(0)}, \alpha^{(1)}, \nu^{(1)}_x, \pi)$, since there are no closed-form solutions. Estimates for parameters of the joint distribution of X and Y for the respondents $(\alpha^{(0)}, \nu^{(0)}_x, \nu^{(0)}_y, \rho^{(0)})$ are obtained by maximizing the likelihood arising from the distribution function given in (2) for respondents ($M = 0$). The marginal distribution of X for nonrespondents ($M = 1$) is Gamma$(\alpha^{(1)}, \nu^{(1)}_x)$ so we take the estimates of these parameters to be the standard gamma MLEs. The ML estimate for π is the proportion of respondents. In the second step, the MLEs of the remaining parameters $(\nu_y^{(1)}, \rho^{(1)})$ are estimated for a chosen value of λ, by plugging the MLEs of the identifiable parameters into (4) for $\lambda = 0$ and (5) for $\lambda = \infty$.

Once the MLEs for all parameters are obtained, substituting them into (6) yields the ML estimate of the mean of Y. In the normal PPM model, modifications to the parameters identified via restrictions are sometimes needed to ensure that estimates lie within the appropriate sample space [Little (1994)]. Similarly, for the gamma PPM model, under MNAR ($\lambda = \infty$), if $\hat{\alpha}^{(1)}\hat{\nu}_x^{(0)} < \hat{\alpha}^{(0)}\hat{\nu}_x^{(1)}(1 - \hat{\rho}^{(0)})$ then $\hat{\rho}^{(1)}$ will be negative. This may occur when the mean of the proxy for nonrespondents is much smaller than the mean for respondents and the proxy is weak ($\hat{\rho}^{(0)}$ close to zero). In this case, the weak relationship between the proxy and the outcome, coupled with the size differential between respondents and nonrespondents, leaves the PPM model very little information with which to adjust estimates for nonignorable nonresponse (MNAR). If this occurs, $\hat{\rho}^{(1)}$ should be set to zero, and whether the gamma PPM model is the appropriate model for the data should be reevaluated. A large-sample variance estimate for μ_y is available through inversion of the information matrix, but unlike the normal PPM, this variance estimate must be solved numerically since there is no closed-form solution for the ML estimates.

3.2. *Multiple imputation.* A limitation of ML estimation for the gamma PPM model is that it treats the proxy X as known, when in fact it is constructed using estimated regression parameters. An alternative estimation method is multiple imputation [Little and Rubin (2002)], which can incorporate uncertainty in the proxy X by including the regression parameters that create X into the imputation framework. We create K complete data sets for a specified value of λ and estimate the adjusted mean of Y using these data. In what follows we describe the three steps required to perform imputation using the gamma PPM model: (1) generating "draws" of the proxy X, (2) generating draws of the parameters of the gamma PPM model and (3) generating draws for the missing Y.

In the first step, "draws" of the proxy X are generated. We place noninformative priors on the regression coefficients in the standard ordinary least squares regression model (without an intercept) that is used to create the proxy X (i.e., the regression of Y on Z) and draw the parameters of the regression model from their posterior distribution. Using these parameter draws we create a "draw" of the proxy X for respondents and nonrespondents. Since the gamma PPM model requires $Y > 0$, the proxy values must be checked to ensure that they are all positive. In our application, the proxy values are rarely nonpositive. However, if a given draw of the regression parameters results in any negative proxy values, these parameter draws are discarded and new draws are generated until the all proxy values are positive. Only for data with a very weak proxy will this be a major problem, and in this case imputation under the gamma PPM model—or really any linear regression model that includes X—may not be an appropriate choice.

In the second step, the remaining parameters of the gamma PPM model are drawn (conditional on the proxy created in the previous step) using algorithms described by Iliopoulos, Karlis and Ntzoufras (2005) for the bivariate KBGD($\alpha^{(0)}, \nu_x^{(0)}, \nu_y^{(0)}, \rho^{(0)}$) and using standard Bayesian inference for the univariate gamma ($\alpha^{(1)}, \nu_x^{(1)}$). Details can be found in Appendix A.

In the final step, once draws for all parameters are obtained, the missing values of Y are drawn based on the conditional distribution of Y given X for nonrespondents ($M = 1$), which for the KBGD is a randomized gamma distribution of the first type. Draws from this distribution are obtained by first drawing a Poisson random variate, and then, conditional on this variate, drawing a Gamma variate [Makarov and Glew (2010)]. For the gamma PPM model, this results in the following draws to impute Y (where the parameters are their values at the current iteration of the imputation algorithm):

$$W_i \sim \text{Poisson}\left(\frac{\rho^{(1)}}{1-\rho^{(1)}}\nu_x^{(1)} x_i\right),$$

$$Y_i | W_i \sim \text{Gamma}\left(W_i + \alpha^{(1)}, \frac{\nu_y^{(1)}}{1-\rho^{(1)}}\right).$$

For imputation, we allow a burn-in of the Gibbs sampler and thin the chain to reduce auto-correlation between imputations, cycling through until a total of K completed data sets are produced.

For the kth completed data set, standard combining rules are used to estimate the mean of $Y(\hat{\mu}_y)$, and its variance, $V(\hat{\mu}_y)$, which breaks down into between (B) and within (W) components of variance [Little and Rubin (2002)]. The estimate of the mean of Y in the kth completed data set is the weighted (by $\hat{\pi}$) average of the ML estimates from the respondent and nonrespondent gamma distributions. As usual, the between-imputation variance is estimated by the variance of these K mean estimates and the within-imputation variance is the average of the K estimates of the variance of the mean. For the kth completed data set, the estimated variance of the mean is given by the ML estimate of the variance of the mixture of gamma distributions divided by the total sample size.

3.3. *Estimating the fraction of missing information with the gamma PPM model*. The fraction of missing information (FMI) has been proposed as a metric for assessing the risk of nonresponse bias for a specific adjusted survey estimate [Wagner (2012)], and as a measure of survey quality that can be monitored during data collection [Wagner (2010)]. The FMI is a measure of loss of precision due to nonresponse and is the ratio of between-imputation variance to total variance for a specific estimator [Little and Rubin (2002)]. Under the PPM framework, Andridge and Little (2011) propose using the *set* of FMI values obtained in their sensitivity analysis as a marker of the potential for nonresponse bias and the ability to correct potential bias. A useful threshold is the nonresponse rate or the imputation rate for an outcome variable as described in Thompson and Oliver (2012). An FMI below either rate indicates a strong proxy X, and therefore good information with which to correct for bias, even under an MNAR mechanism ($\lambda = \infty$). If the FMI values are close together, then the inflation of variance due to an MNAR mechanism is not severe, relative to the MAR mechanism. However, if the range of FMI values is large, and especially if the largest value approaches the maximum value of 1, then this indicates a lack of information for assessing bias. Of course, if the minimum value of the FMI is close to 1—or both values were exactly equal to 1—then one could conclude that the imputation model used to derive the proxy is inadequate for ameliorating nonresponse bias effects under any response mechanism. For a more detailed discussion of the factors impacting FMI and its use in the PPM framework, see Andridge and Thompson (2015a).

As with the normal PPM model, estimates of FMI can be obtained under the gamma PPM model. The FMI is a natural byproduct of the multiple imputation approach and is estimated as the ratio of the between-imputation variance to the total variance of a specific estimator, the overall mean of Y in our application, with an adjustment factor based on K for the finite number of imputations [Little and Rubin (2002)]. An ML estimate of the FMI can also be obtained from the ML estimates identified via parameter restrictions (see Appendix B), but no closed-form solution exists.

4. Application to the service annual survey (SAS).

4.1. *Empirical data background.* In this section we apply the normal and gamma PPM models to six years of collection data in selected industries from the Service Annual Survey (SAS). We focus on total revenue and total expenses, the key items collected by the SAS; payroll is collected along with expenses. For the SAS, when ratio imputation is used, total payroll is regressed on total expenses, and total expenses are regressed on total revenue. We therefore perform two separate PPM analyses: one using expenses (Z) to predict payroll (Y), and using revenue (Z) to predict expenses (Y). Imputation parameters are developed within industry and tax-status category (taxable and tax-exempt).

Our empirical analysis uses data from three industries:

- Farm Product Warehousing and Storage (NAICS 493130).
- Miscellaneous Ambulatory Health Care Services (NAICS 621999).
- Psychiatric and Substance Abuse Hospitals (NAICS 622218).

Because of the lengthy titles associated with each industry, we use their associated industry codes in the figures and tables below. These industries represent a very small cross-section of the SAS sample and are not meant to represent the larger survey in its entirety. Instead, they were recommended by subject matter experts as candidates for a feasibility study because they present the realistic situations often encountered with business surveys' response patterns (discussed below). This set of industries also allows us to assess differences between the normal and gamma PPM models. Note that our estimates differ from the published estimates. First, we exclude all cases that have missing or zero-value independent variable values, as well as any observations with an imputed zero-valued dependent variable. In practice, zero-value revenue, expenses, and payroll responses are imputed deterministically via separate analysis procedures that depend on the sample unit business classification (e.g., non-employer, not-for-profit). Second, we consider only ratio imputation. In practice, the missing data for the larger units are more likely to be replaced by administrative or other auxiliary data (e.g., values from published reports), which in turn would greatly reduce the effect of imputation error on the totals [Beaumont, Haziza and Bocci (2011)]. Total revenue and total expenses are rarely changed in practice by the "logical edit" imputation, although the reported value of total payroll may be modified slightly since it is a component of total expenses. Our PPM analyses therefore represent "worst case" scenarios for nonresponse bias analyses of the SAS key estimates.

In the Farm Product Warehousing and Storage industry (NAICS 493130), the response rates for each unit size strata are approximately the same so that nonresponse does not appear to depend on size. In the Miscellaneous Ambulatory Health Care Services industry (NAICS 621999), the largest units (the certainty units) respond at a lower rate than the smaller noncertainty units, and among noncertainty units, the respondents tend to be *smaller* than the nonrespondents. The Psychiatric and Substance Abuse Hospitals Industry (NAICS 622218) is the most "typical" of the data sets for a business population; larger (certainty) units respond at a higher rate than the smaller noncertainty units, and among noncertainty units the respondents tend to be larger than the nonrespondents. The median (unweighted) unit nonresponse rate with our research data was 32% across these industries and years.

To assess the appropriateness of the gamma assumption of the PPM model, we performed three separate checks using the SAS data. First, we plotted the empirical CDFs of the predicted proxies within industry against the theoretical CDFs obtained using maximum likelihood estimates for normal and gamma distributions ($\theta = 0$). The empirical CDFs incorporate the sampling weights as recommended in Lohr (2010), and consequently have a pronounced "step pattern" with concentrated mass on smaller units that have large sampling weights. Because of this, goodness-of-fit tests that compared empirical CDFs to the smooth theoretical distributions such as the Kolmogorov–Smirnov test have very poor power and thus were not performed. Figure 2 presents "typical" results from one industry from the 2005 data collection. The complete set of results for all studied years and models are available upon request. In all cases, there is little evidence for a normally distributed proxy, whereas the theoretical gamma distributions appear to better fit the data. As a second validation of the gamma distribution assumption, we simulated

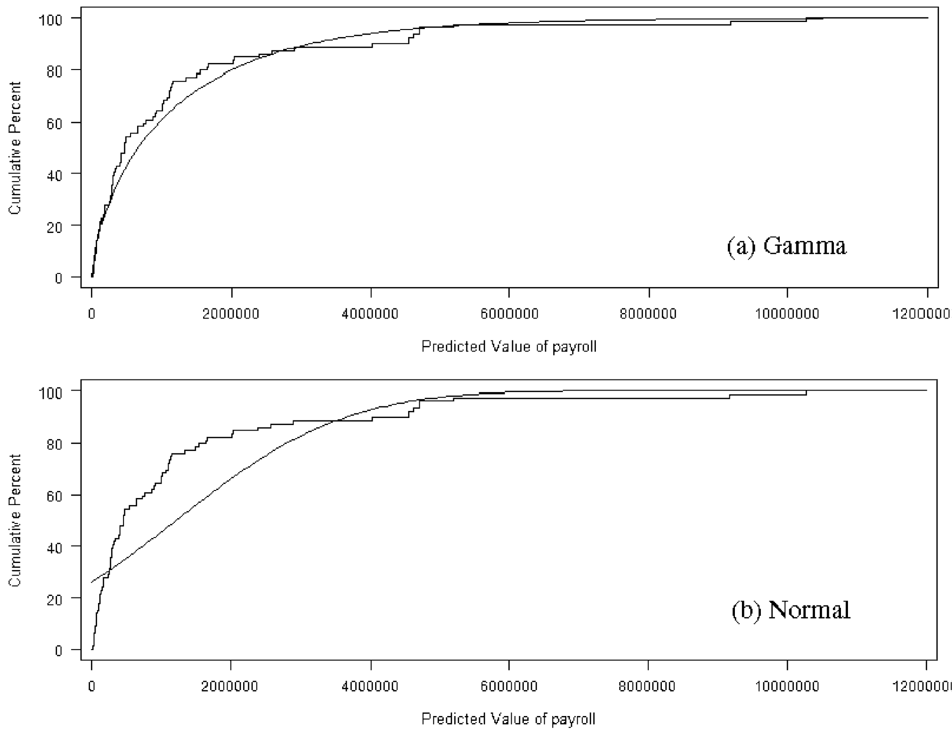

FIG. 2. *Empirical and theoretical CDF for* (a) *gamma distributed proxy and* (b) *normally distributed proxy in the "Farm Product Warehousing and Storage" Industry* (*NAICS* 493130), 2005 *data*, *for the SAS payroll proxy model*.

gamma-distributed populations for our predicted proxy using method-of-moment parameters (with sampling weights) from the empirical data in the studied industries and compared the first three moments from the simulated populations to the (weighted) empirical data moments (sample mean, sample variance and sample skewness). As a third and final check, we verified that the assumption of a shared shape parameter in the gamma PPM was reasonable by estimating the shape parameters of the outcome and estimated proxy among respondents using standard method-of-moments estimators and confirming that they were similar.

We applied multiple imputation using both the normal and gamma PPM models to the three industries' data separately by year, for both payroll and expenses data. We needed a large number of imputed data sets in order to estimate the FMI with reasonable precision [Harel (2007)], and so for the normal multiple imputation (MI) we created $K = 500$ imputed datasets. For the gamma MI, we allowed a burn-in of 500 draws and afterwards thinned the chain by 10 to obtain a total of $K = 200$ imputed data sets; the smaller number of imputations was due to a high computational burden. As a consequence of the smaller number of imputations, we expect that the gamma model FMI estimates will be more subject to more random noise than their normal counterparts. That said, we did examine results obtained with larger numbers of imputations for a subset of industries/years and found that the differences were generally in the third or fourth decimal places, not sufficient to change interpretation. We also performed ML estimation for both models. The main difference between the estimation methods was a slight underestimation of variance by the ML models compared to the MI models, which is as expected, since proxy variability is ignored in the ML estimation. However, the comparison between the gamma and normal models was essentially the same and so we do not report the ML results.

4.2. *Results*. Figure 3 presents mean estimates and 95% confidence intervals from multiple imputation using the gamma PPM model and the normal PPM model for $\lambda = 0$ (MAR) and $\lambda = \infty$ (MNAR) in the three industries for the payroll model. Results for the expenses model were similar and are available in the supplemental material [Andridge and Thompson (2015b)]. Confidence intervals were constructed as described in Little and Rubin (2002) for multiply imputed data using a t reference distribution. As expected, within the same response mechanism (MAR or MNAR), the estimated means ($\hat{\mu}_y$) from the normal and gamma models are generally very close, and each set of means shifts downward from $\lambda = 0$ (MAR) to $\lambda = \infty$ (MNAR). Thus if we were simply interested in estimating mean payroll (or expenses) for the SAS data, either the normal or the gamma model could be used, and the same conclusions drawn. In general, the mean shift from MAR to MNAR within a model is larger than the difference in means between the gamma and normal models within a response mechanism, though the shift is very small in magnitude in some cases (e.g., NAICS 493130, 2005 data). On average across the industries and years, the MNAR mean is 3% smaller than the MAR mean. Consequently, if data were truly MNAR (an unverifiable assumption), these results suggest that the imputation procedures that assume MAR produce overestimates of means. The magnitude of the shift from MAR to MNAR is related to the strength of the proxy. For example, there is a more pronounced shift in the means from MAR to MNAR for the Farm Product Warehousing and Storage industry and the Miscellaneous Ambulatory Health Care Services industry (NAICS 493130 and 621999), which have the weakest proxies, compared to the shift in the means in the Psychiatric and Other Substance Abuse Hospitals industry (NAICS 622218), which is barely noticeable due to the extremely strong proxies. However, in all cases the shift in means is quite small relative to the size of the confidence intervals; in this case we would conclude that the penalty for assuming MAR (as is done by the standard SAS imputation procedures) if data were in fact MNAR is small relative to the overall variability in the mean estimates.

Although the means are similar for the normal and gamma PPM models, the variances are quite different. Variance estimates from the normal model were larger than the gamma model for the payroll data across all but one 18 industry/year combinations under MAR and across all combinations under MNAR. The average

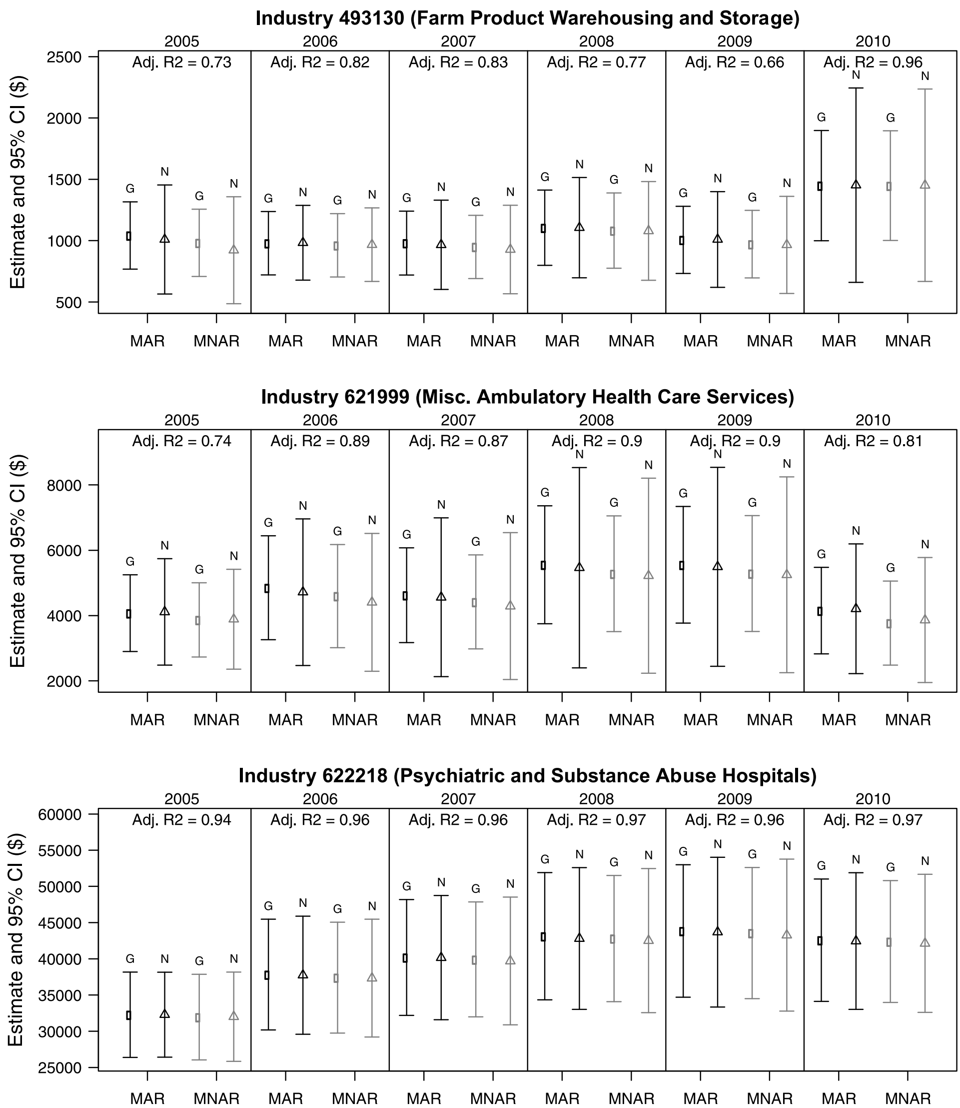

FIG. 3. *Empirical mean and* 95% *confidence intervals for the SAS payroll proxy model for three industries. MAR* = *Missing at Random* ($\lambda = 0$); *MNAR* = *Missing Not at Random* ($\lambda = \infty$). G = *Gamma model*; N = *Normal model.*

overestimation was 93% for the MAR models and 89% for the MNAR models. Note that the scale of the y-axis is different in each industry, and the apparently small differences in confidence interval widths in the Psychiatric and Other Substance Abuse Hospitals industry (NAICS) 622218 are about the same magnitude as in the other industries.

TABLE 1
Multiple imputation estimates of FMI using the normal and gamma payroll proxy models, for both MAR ($\lambda = 0$) *and MNAR* ($\lambda = \infty$), *for three SAS industries*. R^2 *values are the median value across years for each industry*

		Data collection (Year)											
		2005		**2006**		**2007**		**2008**		**2009**		**2010**	
Industry	**Model**	$\lambda = 0$	$\lambda = \infty$	$\lambda = 0$	$\lambda = \infty$	$\lambda = 0$	$\lambda = \infty$	$\lambda = 0$	$\lambda = \infty$	$\lambda = 0$	$\lambda = \infty$	$\lambda = 0$	$\lambda = \infty$
493130	Normal	0.13	0.18	0.06	0.09	0.10	0.12	0.11	0.18	0.18	0.34	0.03	0.04
($R^2 = 0.80$)	Gamma	0.04	0.06	0.05	0.07	0.07	0.09	0.11	0.14	0.11	0.16	0.04	0.03
621999	Normal	0.28	0.38	0.16	0.17	0.19	0.22	0.12	0.14	0.12	0.14	0.27	0.36
($R^2 = 0.88$)	Gamma	0.16	0.17	0.04	0.04	0.06	0.07	0.07	0.08	0.05	0.08	0.14	0.19
622218	Normal	0.10	0.11	0.03	0.03	0.03	0.03	0.03	0.03	0.04	0.04	0.03	0.03
($R^2 = 0.96$)	Gamma	0.12	0.15	0.01	0.01	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.02

The effect of the differences in estimated variances manifests itself in the FMI estimates, which is the metric we proposed using in Section 1 to assess the risk for nonresponse bias in the SAS. Table 1 presents the multiple imputation-based estimates of FMI along with median R^2 using both the normal and gamma PPM models for the payroll data. Often the differences in corresponding FMI for normal and gamma PPM models are not trivial. The first two industries (493130, 621999) have the weakest proxies, and the Miscellaneous Ambulatory Health Care Services industry (621999) has the highest nonresponse rates. Here, the normal FMI values are much larger than the gamma FMI counterparts in most cases, demonstrating the effect of model misspecification when there is a nonnegligible missingness rate coupled with a weaker proxy. For these two industries, if we use the normal model, the larger FMI values provide evidence *against* the subject matter experts' contentions that their imputation procedures have mitigated the effects of nonresponse on key survey estimates. In contrast, the smaller gamma model FMIs are more in line with the "commonly held wisdom" that the SAS imputation procedures are reducing the impact of nonresponse.

The Psychiatric and Other Substance Abuse Hospitals (NAICS 622218) represents the most "typical" response pattern of the three studied industries; respondents are larger on average than nonrespondents, so that response propensity is related to unit size. However, this industry also has the smallest sample size (median $n = 73$) and the strongest proxies ($R^2 \geq 0.92$ for all years). Looking at the FMI estimates in Table 1 shows that for this industry the normal FMI estimates tend to be approximately equal to or slightly smaller than their gamma counterparts, which is the reverse of what is seen with the other two industries (493130, 621999). In this case, there is little difference in the interpretation of the effect of nonresponse bias if using a misspecified model.

Notice that in two instances the gamma FMIs for $\lambda = 0$ are slightly larger than their $\lambda = \infty$ counterparts (2010, industries 493130 and 622218). For both of these cases the proxy model is very strong (R^2 values of 0.96 and 0.97, respectively) and the proxy means for respondents and nonrespondents are close together, leading to FMI estimates very close to zero for both $\lambda = 0$ and $\lambda = \infty$. We believe that these FMI estimates (for $\lambda = 0$ and ∞) are not significantly different, and it is just the finite number of imputations causing the small difference in the unexpected direction.

The contrast between normal and gamma results supports the theory presented in Section 3. Within the same response mechanism (MAR or MNAR), the normal and gamma models produce similar mean estimates, but variance estimates (and thus FMI estimates) can be substantially different, even with very strong auxiliary information (high R^2 values). In the SAS data we have very strong predictors of revenue and expenses from other survey items as demonstrated here; these high model R^2 values for the key items are not atypical. However, not all applications will have such strong proxy variables. In fact, other items collected by the SAS such as the detailed revenue or detailed expenses subcomponent items have much weaker relationships with candidate predictors. We expect the disagreement between the models to be more severe with weaker proxies and with higher rates of nonresponse, making choosing the appropriate model even more important. To investigate this further we performed a simulation study.

5. Simulation study. We conducted a small simulation study to explore the contribution of proxy strength and differing sizes (and variances) of respondents and nonrespondents to inference obtained under the normal and gamma PPM models. We expected both models to produce approximately unbiased mean estimates, as we saw similar mean estimates for both models with the SAS data. However, we expected the normal model to overestimate (underestimate) variances when small (large) Y were missing. Correspondingly, we also expected to see the normal methods to overestimate (underestimate) the FMI when small (large) Y were missing.

5.1. *Data generation.* The outcome Y and a single covariate Z were first generated from the KBGD($\alpha, \nu_y, \nu_z, \rho$) for a sample of size n, using the method of Ong (1992) to generate the draws. To reduce the size of the experiment while focusing attention on key parameters, we fixed the total sample size at $n = 100$, which was approximately the median sample size for the SAS industries studied. We fixed the shape parameter at $\alpha = 1$ and the rate parameters at $\nu_y = 0.02$, $\nu_z = 0.01$. These shape and rate parameters were based on method of moment estimates from empirical analysis of the studied SAS datasets, though rate parameters were scaled to be more tractable for the simulation (preserving the shape of distributions). We considered three different values for the correlation, $\rho \in \{0.5, 0.7, 0.9\}$. In our studied SAS data sets, the covariates are highly predictive of the outcome, with R^2 values above 0.75. However, this is not always the case in other data sets (or even with other survey items in the SAS), so we included lower correlation values to evaluate differences in methods with weaker proxies. A total of 500 replicates of each parameter combination were used.

The missing data indicator M was generated via random draws from the Bernoulli distribution with probability according to a logistic regression model,

$$\text{logit}(\Pr(M = 1|Y, Z)) = \gamma_0 + \gamma_Z Z + \gamma_Y Y,$$

and values of Y were deleted when $M = 1$. We considered both MAR and MNAR mechanisms, as well as varied whether respondents or nonrespondents were larger on average. The four missing data scenarios and corresponding parameter values for the logistic model were: (1) MAR with nonrespondents larger than respondents ($\gamma_0 = -1$; $\gamma_Z = 0.01$; $\gamma_Y = 0$); (2) MAR with nonrespondents smaller than respondents ($\gamma_0 = 1$; $\gamma_Z = -0.01$; $\gamma_Y = 0$); (3) MNAR with nonrespondents larger than respondents ($\gamma_0 = -1$; $\gamma_Z = 0$; $\gamma_Y = 0.02$); (4) MNAR with nonrespondents smaller than respondents ($\gamma_0 = 1$; $\gamma_Z = 0$; $\gamma_Y = -0.02$). The values of γ_0 were chosen to induce a 50% nonresponse rate, which was the highest nonresponse rate in all of the studied SAS data sets and represented our "worst-case" scenario.

We note that in our simulations, the gamma PPM is not the data generation model, since the gamma PPM implies the joint distribution of Y and X is KBGD *conditional* on M, whereas in the simulation the *unconditional* joint distribution of Y and Z is the KBGD. In this simulation the distributional assumptions of the gamma PPM model are therefore violated, so that in addition to comparing the normal and gamma models, we can assess how well the gamma PPM model performs when it is not exactly the data generation model. All steps of the simulation study were conducted using R [R Core Team (2012)].

5.2. *Estimation methods*. For each replicate, complete data were generated, missingness was imposed, and multiple imputation was performed under the gamma PPM model and the normal PPM model. As with the SAS application, we also performed estimation using maximum likelihood; variance estimates were slightly smaller (within a model) as expected, but overall conclusions comparing the normal and gamma models were similar as with MI and thus are not discussed further. For scenarios 1 and 2 we used $\lambda = 0$ for the PPM models, corresponding to an assumption of MAR, and for scenarios 3 and 4 we used $\lambda = \infty$, corresponding to an assumption of MNAR. In order to estimate the FMI with reasonable precision, we used a total of 200 imputed data sets for both the normal and gamma MI procedures. The normal MI is noniterative; for the gamma MI, we allowed a burn-in of 500 draws and then imputed on every tenth draw. This resulted in little to no correlation between parameter draws used for imputation for all parameters except $\rho^{(0)}$, for which a low level of autocorrelation remained. A higher thinning value would be desirable but was not possible due to the computational intensity of the procedure; results were averaged over 500 simulation replicates and thus are still reasonable.

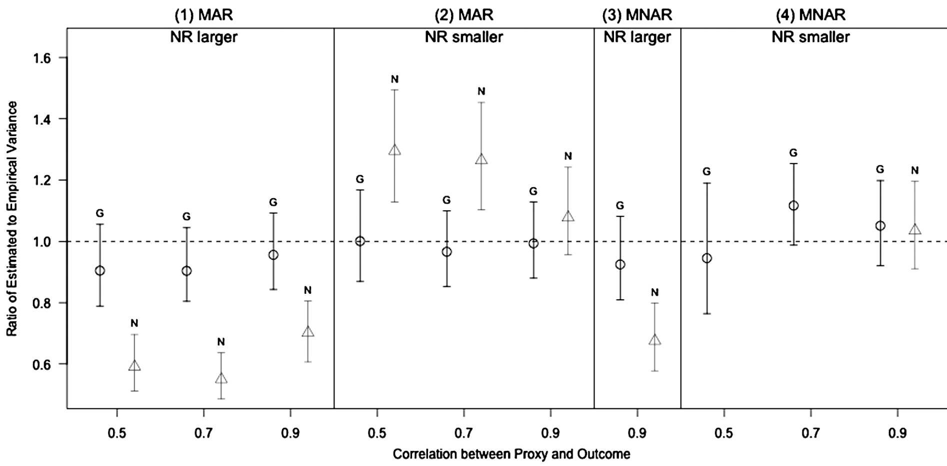

FIG. 4. *Ratios of median estimated variance to empirical variance estimate from the simulation study for the gamma and normal PPM imputation models*, *with* 95% *bootstrap confidence intervals* (1000 *bootstrap samples*). *Numbers* 1–4 *refer to the four simulation scenarios*. *Results over* 500 *replicates*. MAR = *Missing at Random* ($\lambda = 0$); MNAR = *Missing Not at Random* ($\lambda = \infty$); NR = *Nonrespondents*; G = *Gamma model*; N = *Normal model*.

To illustrate how well each method estimated the mean of Y, we calculated the median relative bias of the estimate of μ_y, noting that the true mean of Y was 50 across all scenarios. Since we were especially interested in assessing estimation of the variance of $\hat{\mu}_y$ we calculated the ratio of the median estimated variance of $\hat{\mu}_y$ to the empirical variance of $\hat{\mu}_y$. Estimates of the simulation error for this ratio were obtained using the bootstrap [Efron (1994)], since each set of 500 replicates only provided a single point estimate of the ratio. One thousand bootstrap samples were drawn from the set of 500 mean and variance estimates, and for each bootstrap sample the ratio was recalculated; 2.5th to 97.5th percentiles of the resulting bootstrap distribution of the ratio are provided together with the point estimates (see Figure 4). We also computed the actual coverage of a nominal 95% interval. The median FMI for each method across the 500 simulation replicates is reported in order to compare models; we note that the true value of FMI is not easily calculated. Medians are reported instead of means to reduce the impact of occasional very large imputed values that occurred with lower correlations under the MNAR mechanism, inflating mean and variance estimates.

5.3. *Results*. For the normal MI model, when smaller units were missing and data were MNAR, parameter draws frequently got stuck trying to find draws that produced nonnegative variance parameters. This will happen for the normal model when the respondent variance is much smaller than the nonrespondent variance. This is of particular concern for the SAS applications (and business survey applications in general) since survey respondents tend to be the larger sampled units

TABLE 2
Median empirical relative bias, coverage, and fraction of missing information (FMI) for the simulation study. Results over 500 replicates. MAR = Missing at Random, MNAR = Missing Not at Random. Bolded coverages are below 1.96 simulation standard errors; shaded cells are coverages above 1.96 simulation standard errors

Scenario			ρ	Model	Relative bias (%)	Coverage	FMI
1	MAR ($\lambda = 0$)	Nonrespondents larger	0.5	Gamma	−2.6	**0.90**	0.56
				Normal	−0.7	**0.86**	0.52
			0.7	Gamma	−3.1	**0.92**	0.49
				Normal	−1.7	**0.86**	0.40
			0.9	Gamma	−2.7	0.94	0.27
				Normal	−2.3	**0.89**	0.18
2	MAR ($\lambda = 0$)	Nonrespondents smaller	0.5	Gamma	1.5	0.96	0.40
				Normal	−1.2	0.97	0.51
			0.7	Gamma	−0.2	0.95	0.28
				Normal	−2.3	0.96	0.44
			0.9	Gamma	−0.8	0.95	0.11
				Normal	−2.1	0.96	0.25
3	MNAR ($\lambda = \infty$)	Nonrespondents larger	0.5	Gamma	16.2	0.98	0.92
				Normal	2.0	0.94	0.82
			0.7	Gamma	5.4	0.95	0.71
				Normal	1.6	**0.91**	0.55
			0.9	Gamma	−1.1	0.94	0.32
				Normal	−2.9	**0.90**	0.20
4	MNAR ($\lambda = \infty$)	Nonrespondents smaller	0.5	Gamma	0.7	0.97	0.68
				Normal	−*	–	–
			0.7	Gamma	−2.9	0.94	0.45
				Normal	–	–	–
			0.9	Gamma	−2.4	**0.93**	0.13
				Normal	−2.1	0.94	0.30

*The Normal MI model did not run for MNAR with nonrespondents smaller and lower correlations ($\rho = 0.5, 0.7$).

and it is impossible to assess the cause of unit nonresponse (e.g., are certain units not responding because the collection is burdensome?) without successfully debriefing the nonrespondents. As a result, no results were obtained for the normal MI model when data were MNAR with nonrespondents smaller than respondents (scenario 4) and $\rho \in \{0.5, 0.7\}$.

When data were MAR, both the normal and gamma models were either unbiased or slightly underestimated the mean (Table 2), and the relative bias was no larger than 3.1% for either model in any MAR scenario. When data were MNAR, the normal MI model had low relative bias when it could provide estimates. The gamma MI model had low relative bias everywhere except when large units were missing and the correlation was low ($\rho = 0.5$). In this case, the gamma MI procedure occasionally imputed extremely large values, skewing the MI means and inflating variance estimates. In practice, one could discard these draws, but in the simulation study they skewed the results and thus the median relative bias across the simulation replicates was relatively high for this case.

More striking differences between the two models arose when looking at the estimated to empirical variance ratio, as well as coverage and the FMI. Figure 4 shows the ratio of estimated to empirical variance along with 95% bootstrap intervals to reflect simulation error. When data were MNAR and large units were missing (scenario 3), the occasional extremely large imputed values for lower correlations under both models made comparing the estimated and empirical variance uninformative, since both values were drastically inflated, thus the variance ratio is not shown. In these cases coverage was at or above nominal for the gamma model and there was slight undercoverage for the normal model for these situations (Table 2).

The normal model underestimated the variance of when larger units were missing, with estimated variance as much as 45% below the empirical variance for the MI method. Coverage was therefore below nominal for all scenarios (Table 2). Conversely, the normal model overestimated the variance of when smaller Y were missing, with estimated variance as much as 30% above the empirical variance for MI. The poor performance of the normal models was somewhat worse for weaker proxies (smaller ρ). In contrast, for the gamma model, estimated variances were closer to empirical variances (within simulation error) for all scenarios. We note that the biased variance estimates of the normal PPM model are clearly illustrated with the variance ratios in Figure 4, but are somewhat less evident looking at the coverage values in Table 2, especially for the scenarios where there is overestimation. Since our initial concern with the normal PPM was its ability to produce unbiased estimates of variances, the variance ratios were a key component of the simulation.

Table 2 also shows the FMI estimates for all scenarios for each model. As expected, FMI estimates from the gamma model were larger than from the normal model when larger units were missing, and were smaller when smaller units were missing. These differences were relatively constant across the different correlations, but were more exaggerated when smaller units were missing. In some cases the differences between the gamma and normal model estimates of FMI were quite striking. For example, when data were MAR with small respondents missing and a high correlation of $\rho = 0.9$, the FMI estimate for the gamma model was 11% compared to 25% for the normal model, a greater than two-fold difference. We note that the FMI values in the simulation are quite a bit larger than those seen in the SAS data; this is due to the simulation data having a higher rate of nonresponse and weaker proxies than most of the SAS data sets.

The results of the simulation study illustrate that the normal PPM model is robust to departures from normality when estimating means, but not necessarily when estimating variances. In particular, with positively skewed data as seen in the SAS data, the normal PPM model can produce highly biased variance estimates. The direction of the bias depends on the direction of differences between respondents and nonrespondents in terms of the proxy variable. If nonrespondents tend to be larger than respondents then the normal model will underestimate variances, and if nonrespondents tend to be smaller then it will overestimate variances. Because the variance estimates from the normal PPM model are biased, the FMI estimates are likewise affected and can be misleading. In contrast, the gamma PPM model that we developed specifically for use with the skewed SAS data does not have the same deficiencies in variance estimation, as demonstrated in the simulation where estimated to empirical variances are close to one (despite the data generation model not being the gamma PPM model). Thus in the SAS applications, where the skewed data are well represented by gamma distributions, when we saw differences between the two models' FMI values, we give more credibility to results from the gamma PPM model over the normal PPM model. The distinctions between the two sets of FMI values are particularly important in the case of a "weak" proxy and a nonnegligible missingness rate (industries 493130 and 621999), where the normal FMI estimates tend to be about twice as large as their gamma counterparts. If we were to "believe" the normal model results (we do not), then these results would provide evidence against the subject matter experts' contention that the item-specific imputation procedures offset nonresponse bias in the key estimates. However, results from the gamma model, which better fits the SAS data, support their contention.

6. Conclusion. Federal surveys are required to perform nonresponse bias analyses when either their unit response rate or key item response rates fall below performance benchmarks. At a minimum, such analyses need to determine the existence of nonresponse bias in survey estimates and to measure, when possible, the effects of these biases. Within the same program, the effect of nonresponse bias—if it exists—may differ by item. Consequently, different mitigation strategies may be needed within the same survey.

The PPM analysis proposed by Andridge and Little (2011) provides an item-specific method of evaluating the effects of nonresponse bias on survey data. The usage of the FMI as objective criteria is quite appealing for a survey manager as it is easily interpretable and can be computed before and after mitigation strategies such as new imputation or weighting strategies introduced in order to assess their effectiveness (a decrease in the FMI would indicate an improvement).

However, the originally proposed PPM methodology assumes that the underlying population distribution is multivariate normal. This assumption is often quite reasonable for a demographic survey. However, business survey populations like the SAS often have highly positively skewed characteristics. Thus we developed a PPM model that uses a bivariate gamma distribution. The empirical results from the SAS presented in Section 4 demonstrate situations where results from the normal model would be quite different from those obtained under the gamma model and could lead to erroneous conclusions, particularly as they relate to variance and FMI estimates. In the presented analyses, the normal model FMIs tend to be much larger than their gamma model counterparts in two of the three industries, although the range of FMIs (under differing response mechanisms) is about the same. From a survey manager's perspective, the distinction between the two sets of FMIs is important. Recall that the SAS managers contend that their survey's imputation procedures essentially eliminate the effect of nonresponse bias on the adjusted estimates. The gamma model FMIs support this claim, whereas the normal model FMIs are less confirmatory.

The gamma PPM model has attractive features. The form of the bivariate gamma distribution used was easily adaptable to the PPM framework, since both the conditional mean and variance had regression-like forms. This allowed the "inversion" of the model using parameter restrictions to identify nonrespondent parameters when missingness depended on $Y(\lambda = \infty$, MNAR). Additionally, the sensitivity analysis based on λ draws a picture of the potential for nonresponse bias and our ability to correct it under two different assumptions on the missing data mechanisms (MAR, MNAR).

We considered alternate models for the highly skewed and heteroskedastic business data, such as lognormal distributions or imposing normality on the ratio imputation model. However, neither of these methods easily fit into the PPM framework. In the bivariate lognormal distribution neither the conditional mean $E(Y|X)$ nor variance $V(Y|X)$ is linear in X, which was a property we sought given that the ratio imputation model has this form. Another possibility was extending the ratio imputation model by adding a normality assumption to the variance. However, the corresponding joint distribution did not factor into a form amenable to imposing parameter restrictions when missingness depended on Y ($\lambda = \infty$, MNAR). In other words, this model could be used for inference when $\lambda = 0$, but did not provide a method for inference when $\lambda = \infty$.

There are implementation challenges with the gamma PPM model that are not present with the more tractable normal model implementation. First, there is no closed-form solution for ML estimates and numerical solutions are required. Second, while the Bayesian implementation is attractive because it treats the proxy values as estimates rather than as fixed, it requires a large number of iterations, along with a substantial burn-in. Additionally, the algorithm used for the imputations treats gamma scale parameters as fixed (at their ML estimates), which may cause underestimation of variances.

In the simulation study, the multiple imputation approach occasionally produced extremely large draws under the MNAR mechanism ($\lambda = \infty$). This can occur when there is a low correlation between Y and Z, for example, for the scenario where the average correlation was 0.5, some replicates had a correlation very close to zero. In practice, we recommended restricting the multiple imputation procedure to analysis of key survey items, which are expected to have strong predictors (at a minimum, frame variables) and should be fairly well reported. The maximum likelihood approach is available in the case of computation failure and has proven quite useful in other applications. For example, see Andridge and Thompson (2015a) for discussions on the usage of the ML FMI in the selection of variables for regression imputation models or for the development of imputation cells.

Our empirical results using the SAS data illustrate the importance of conducting model validation before applying a PPM analysis. These business survey populations have characteristics that can be well approximated by a gamma distribution. Using the normal PPM model can provide dramatically different results from those obtained with the more appropriate gamma PPM model, and the direction of the difference depends on the characteristics of the response set. Moreover, even if the response mechanism is covariate-dependent, there will be missing values throughout the distribution. Consequently, it is difficult to predict whether the model misspecification will result in over or underestimation of variances.

The procedure that we present is a parametric analysis of survey data. However, we propose using it on populations that cannot be well approximated with normal models, replacing the more tractable distribution with data-appropriate models. This approach is similar to the adaptations for complex survey data to chi-squared tests or logistic regression models proposed by Rao and Scott (1992) and Roberts, Rao and Kumar (1987), that is, determine the appropriate analysis, then modify the test statistic so that it has the correct asymptotic properties given the finite population and sample design. In a similar vein, the MI approach provides an alternative yet comparable measure that can be used to analyze data from small samples when large sampling theory fails; cf. the Fay–Herriot model in small area estimation [Fay and Herriot (1979), Rao (2003)].

The analyses that we conducted did not use survey weights in any step. In our application to the SAS data, we implicitly incorporated the sampling design by conducting each PPM analysis separately within imputation cell (industry by tax-exempt status) and by using predictors that are strongly correlated to unit size. We would have obtained identical results by including indicator variables for strata and for tax-exempt status in these particular models; including the sampling weight additionally would have made little or no difference. Other practitioners should not discard the usage of survey weights in their proxy models without having validating model diagnostics. Survey weights can be incorporated into both PPM models as predictors in the regression model that creates the proxy, and could also be used in the post-MI inference for both gamma and normal models (e.g., use a Horvitz–Thompson estimator).

Ultimately, we believe that the gamma PPM model can provide a concise picture of the nonresponse problem for the SAS and other business surveys, especially as it relates to comparing across missingness assumptions (MAR $\lambda = 0$ vs MNAR $\lambda = \infty$) and across industries (subpopulations) and years. Likewise, the gamma PPM model could be applied to other types of establishment surveys that exhibit skewness, heteroscedasticity and similar patterns of nonresponse such as government surveys or establishment surveys. In our empirical analysis of the SAS data, the level of the FMI was quite low, supporting the appropriateness of the ratio imputation procedures used for these variables and indicating that even if data were missing not at random, strong auxiliary information exists to correct nonresponse bias.

APPENDIX A: BAYESIAN ESTIMATION OF THE GAMMA PPM MODEL

Bayesian estimation of the parameters of the gamma proxy pattern-mixture model proceeds as follows. After drawing the regression parameters that create the proxy and (re-)creating X, the remaining parameters are estimated using algorithms described by Iliopoulos, Karlis and Ntzoufras (2005) for the bivariate KBGD$(\alpha^{(0)}, \nu_x^{(0)}, \nu_y^{(0)}, \rho^{(0)})$ and using standard Bayesian inference for the univariate gamma $(\alpha^{(1)}, \nu_x^{(1)})$.

The shape parameter α is treated as fixed in Iliopoulos, Karlis and Ntzoufras (2005), so we treat both $\alpha^{(0)}$ and $\alpha^{(1)}$ as fixed in our model by setting them equal to the maximum likelihood estimates. As a consequence, variances may be slightly underestimated due to ignoring the variability in these parameters. As in Iliopoulos, Karlis and Ntzoufras (2005), we reparameterize the KBGD, taking $\theta_j^{(0)} = \nu_j^{(0)}/(1 - \rho^{(0)})$ for $j = x, y$ and specify the following prior distributions:

$$\begin{aligned} \theta_j^{(0)} &\sim \text{Gamma}(0.001, 0.001) \qquad \text{for } j = x, y, \\ \rho^{(0)} &\sim \text{Beta}(1, 1), \\ f(\nu_x^{(1)}) &\propto 1/\nu_x^{(1)}. \end{aligned}$$

The posterior distribution for $\nu_x^{(1)}$ is $\nu_x^{(1)} \sim \text{Gamma}((n-r)\alpha^{(1)}, \sum_{i \in NR} x_i)$. For $\theta_x^{(0)}$, $\theta_y^{(0)}$, and $\rho^{(0)}$, the data augmentation approach of Iliopoulos, Karlis and Ntzoufras (2005) is used to obtain draws from their posterior distributions. Briefly, the algorithm augments with an unobserved sample from the negative binomial distribution, $\boldsymbol{\kappa} \sim \text{NB}(\alpha^{(0)}, 1 - \rho^{(0)})$ such that X_i and Y_i are independent conditional on κ_i. These $\boldsymbol{\kappa}$ then have a Bessel distribution conditional on $\theta_x^{(0)}$, $\theta_y^{(0)}$, $\rho^{(0)}$, X_i and Y_i. This results in the posterior distributions of $\theta_x^{(0)}$, $\theta_y^{(0)}$ and $\rho^{(0)}$ being independent Gamma$(\theta_x^{(0)}, \theta_y^{(0)})$ and Beta$(\rho^{(0)})$ distribution, conditional on $\boldsymbol{\kappa}$, X_i and Y_i [see Iliopoulos, Karlis and Ntzoufras (2005) for details] and a Gibbs sampler can be used to sequentially draw from the posteriors. Simulation from the Gamma and Beta distributions is straightforward; random variates from the Bessel distribution are created using the second rejection sampling algorithm of Devroye (2002).

The draws of $\theta_x^{(0)}$ and $\theta_y^{(0)}$ are then back-transformed to obtain draws of $\nu_x^{(0)}$ and $\nu_y^{(0)}$. Draws of $\rho^{(1)}$ and $\nu_y^{(1)}$ are obtained through the transformations given in

(4) and (5), depending on the selected value of the sensitivity parameter λ. Care must be taken to ensure that parameter constraints on $\nu_y^{(1)}$ and $\rho^{(1)}$ are satisfied under MNAR ($\lambda = \infty$). In this case, the drawn value of $\theta_x^{(0)}$ must be larger than $\nu_x^{(1)}\alpha^{(0)}/\alpha^{(1)}$; if this does not hold then $\theta_x^{(0)}$ is redrawn until this condition is met. This check of parameter constraints is not unique to the gamma PPM model; similar restrictions must also be checked for Bayesian implementations of the normal pattern-mixture model [Little (1994)].

APPENDIX B: MAXIMUM LIKELIHOOD ESTIMATION OF THE FMI WITH THE GAMMA PPM MODEL

A maximum likelihood estimate of the fraction of missing information (FMI) for the mean of $Y(\mu_y)$ can be obtained as follows. As described in Section 3.1, the ML estimate of the (total) variance of μ_y is obtained through inversion of the information matrix and must be solved numerically. This produces an estimate of the total variance component of the FMI. The estimate of the within-imputation component is given by

$$\hat{W} = \frac{1}{n}\left[\hat{\pi}\frac{\hat{\alpha}^{(0)}}{(\hat{\nu}_y^{(0)})^2} + (1-\hat{\pi})\frac{\hat{\alpha}^{(1)}}{(\hat{\nu}_y^{(1)})^2} + \hat{\pi}(1-\hat{\pi})\left(\frac{\hat{\alpha}^{(0)}}{\hat{\nu}_y^{(0)}} - \frac{\hat{\alpha}^{(1)}}{\hat{\nu}_y^{(1)}}\right)^2\right]$$

using the ML estimates of the parameters obtained as described in Section 3.1 for a specified value of λ. This expression derives from the variance of a mixture of the two gamma distributions ($M = 0$, $M = 1$). The between-imputation variance component is then estimated by subtracting this within-variance component from the total variance estimate. Thus the FMI can be estimated as the ratio of the between-imputation variance estimate to the total variance estimate.

Acknowledgments. The authors would like to thank Roderick Little, Xijian Liu, Yves Thibadeau, Katrina Washington, three referees, an Associate Editor, and the Editor for their valuable contributions to earlier versions of the manuscript. We also thank Stephen Kaputa for his many contributions to the empirical analysis.

Any views expressed in this paper are those of the author(s) and not necessarily those of the U.S. Census Bureau.

SUPPLEMENTARY MATERIAL

Supplement to "Assessing nonresponse bias in a business survey: Proxy pattern-mixture analysis for skewed data" (DOI: 10.1214/15-AOAS878SUPP; .pdf). The supplementary material contains the results of applying multiple imputation using the gamma PPM model and the normal PPM model for $\lambda = 0$ (MAR) and $\lambda = \infty$ (MNAR) in the three SAS industries for the expenses model.

REFERENCES

ANDRIDGE, R. R. and LITTLE, R. J. A. (2011). Proxy pattern-mixture analysis for survey nonresponse. *Journal of Official Statistics* **27** 153–180.

ANDRIDGE, R. R. and THOMPSON, K. J. (2015a). Using the fraction of missing information to identify auxiliary variables for imputation procedures via proxy pattern-mixture models. *Int. Stat. Rev.* **83** 472–492.

ANDRIDGE, R. R. and THOMPSON, K. J. (2015b). Supplement to "Assessing nonresponse bias in a business survey: Proxy pattern-mixture analysis for skewed data." DOI:10.1214/15-AOAS878.

BAVDAŽ, M. (2010). The multidimensional integral business survey response model. *Survey Methodology* **1** 81–93.

BEAUMONT, J.-F., HAZIZA, D. and BOCCI, C. (2011). On variance estimation under auxiliary value imputation in sample surveys. *Statist. Sinica* **21** 515–537. MR2829845

DEVROYE, L. (2002). Simulating Bessel random variables. *Statist. Probab. Lett.* **57** 249–257. MR1912083

EFRON, B. (1994). Missing data, imputation, and the bootstrap. *J. Amer. Statist. Assoc.* **89** 463–479. MR1294072

FAY, R. E. III and HERRIOT, R. A. (1979). Estimates of income for small places: An application of James–Stein procedures to census data. *J. Amer. Statist. Assoc.* **74** 269–277. MR0548019

FELLER, W. (1966). *An Introduction to Probability Theory and Its Applications. Vol. II*. Wiley, New York. MR0210154

HAREL, O. (2007). Inferences on missing information under multiple imputation and two-stage multiple imputation. *Stat. Methodol.* **4** 75–89. MR2339010

HAZIZA, D., THOMPSON, K. J. and YUNG, W. (2010). The effect of nonresponse adjustments on variance estimation. *Survey Methodology* **36** 35–43.

ILIOPOULOS, G., KARLIS, D. and NTZOUFRAS, I. (2005). Bayesian estimation in Kibble's bivariate gamma distribution. *Canad. J. Statist.* **33** 571–589. MR2232381

IZAWA, T. (1965). Two or multi-dimensional gamma-type distribution and its application to rainfall data. *Papers in Meteorology and Geophysics* **15** 167–200.

KIBBLE, W. F. (1941). A two-variate gamma type distribution. *Sankhyā* **5** 137–150. MR0007218

KREUTER, F., OLSON, K., WAGNER, J., YAN, T., EZZATI-RICE, T. M., CASAS-CORDERO, C., LEMAY, M., PEYTCHEV, A., GROVES, R. M. and RAGHUNATHAN, T. E. (2010). Using proxy measures and other correlates of survey outcomes to adjust for non-response: Examples from multiple surveys. *J. Roy. Statist. Soc. Ser. A* **173** 389–407. MR2751883

KREWSKI, D. and RAO, J. N. K. (1981). Inference from stratified samples: Properties of the linearization, jackknife and balanced repeated replication methods. *Ann. Statist.* **9** 1010–1019. MR0628756

LITTLE, R. J. A. (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika* **81** 471–483. MR1311091

LITTLE, R. J. A. and RUBIN, D. B. (2002). *Statistical Analysis with Missing Data*, 2nd ed. Wiley, Hoboken, NJ. MR1925014

LOHR, S. L. (2010). *Sampling: Design and Analysis*, 2nd ed. Brooks/Cole, Boston, MA. MR3057878

MAKAROV, R. N. and GLEW, D. (2010). Exact simulation of Bessel diffusions. *Monte Carlo Methods Appl.* **16** 283–306. MR2747817

ONG, S. H. (1992). The computer generation of bivariate binomial variables with given marginals and correlations. *Comm. Statist. Simulation Comput.* **21** 285–299.

PEYTCHEVA, E. and GROVES, R. M. (2009). Using variation in response rates of demographic subgroups as evidence of nonresponse bias in survey estimates. *Journal of Official Statistics* **25** 193–201.

R CORE TEAM (2012). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available at http://www.R-project.org/.

RAO, J. N. K. (2003). *Small Area Estimation*. Wiley, Hoboken, NJ. MR1953089

RAO, J. N. K. and SCOTT, A. J. (1992). A simple method of the analysis of clustered binary data. *Biometrika* **74** 577–585.

ROBERTS, G., RAO, J. N. K. and KUMAR, S. (1987). Logistic regression analysis of sample survey data. *Biometrika* **74** 1–12. MR0885914

ROYALL, R. M. (1992). The model based (prediction) approach to finite population sampling theory. In *Current Issues in Statistical Inference*: *Essays in Honor of D*. *Basu*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **17** 225–240. IMS, Hayward, CA. MR1194420

SNIJKERS, G., HARALDSEN, G., JONES, J. and WILLIMACK, D. K. (2013). *Designing and Conducting Business Surveys*. Wiley, New York.

THOMPSON, K. J. (2005). An empirical investigation into the effects of replicate reweighting on variance estimates for the annual capital expenditures survey. In *Proceedings of the Federal Committee on Statistical Methods Research Conference*. U.S. Office of Management and Budget, Washington, DC.

THOMPSON, K. J. and OLIVER, B. E. (2012). Response rates in business surveys: Going beyond the usual performance measure. *Journal of Official Statistics* **28** 221–237.

THOMPSON, K. J. and WASHINGTON, K. T. (2013). Challenges in the treatment of unit nonresponse for selected business surveys: A case study. *Survey Methods: Insights from the Field*. Retrieved from http://surveyinsights.org/?p=2991.

WAGNER, J. (2010). The fraction of missing information as a tool for monitoring the quality of survey data. *Public Opinion Quarterly* **74** 223–243.

WAGNER, J. (2012). A comparison of alternative indicators for the risk of nonresponse bias. *Public Opinion Quarterly* **76** 555–575.

WILLIMACK, D. K. and NICHOLS, E. (2010). A hybrid response process model for business surveys. *Journal of Official Statistics* **1** 3–24.

YUAN, L. and KALBFLEISCH, J. D. (2000). On the Bessel distribution and related problems. *Ann*. *Inst*. *Statist*. *Math*. **52** 438–447. MR1794244

COLLEGE OF PUBLIC HEALTH
DIVISION OF BIOSTATISTICS
OHIO STATE UNIVERSITY
COLUMBUS, OHIO 43210
USA
E-MAIL: andridge.1@osu.edu

U.S. CENSUS BUREAU
ECONOMIC STATISTICAL METHODS DIVISION
WASHINGTON, DC 20233
USA
E-MAIL: katherine.j.thompson@census.gov