Submission to Econometrics Journal (2019), volume **23**, pp. 1–34.
Article No. First submitted in Dec. 2018

IT Outsourcing and Firm Productivity: Eliminating Bias from Selective Missingness in the Dependent Variable [1]

Christoph Breunig[†], Michael Kummer[‡], Joerg Ohnemus[⋆⋆] and Steffen Viete[⋆]

[†]*Spandauer Straße 1, 10178 Berlin, Germany, Humboldt-Universität zu Berlin*
E-mail: `christoph.breunig@hu-berlin.de`
[‡]*REG 2.25E, Norwich Research Park, Norwich NR4 TJ7, U.K., University of East Anglia, GaTech & ZEW*
E-mail: `m.kummer@uea.ac.uk`
[⋆] *P.O. Box 103443, 68034 Mannheim, Germany, Centre for European Economic Research (ZEW)*
E-mail: `steffen.viete@zew.de`
[⋆⋆] *P.O. Box 103443, 68034 Mannheim, Germany, Centre for European Economic Research (ZEW)*
E-mail: `joerg.ohnemus@zew.de`

Received: ...

Summary

Missing values are a major problem in all econometric applications based on survey data. A standard approach assumes data are *missing-at-random* and uses imputation methods, or even listwise deletion. This approach is justified if item non-response does not depend on the potentially missing variables' realization. However, assuming *missing-at-random* may introduce bias if non-response is, in fact, selective. Relevant applications range from financial or strategic firm-level data to individual-level data on income or privacy-sensitive behaviors.

In this paper, we propose a novel approach to deal with selective item non-response in the model's dependent variable. Our approach is based on instrumental variables that affect selection only through a partially observed outcome variable. In addition, we allow for endogenous regressors. We establish identification of the structural parameter and propose a simple two-step estimation procedure for it. Our estimator is consistent and robust against biases that would prevail when assuming missingness at random. We implement the estimation procedure using firm-level survey data and a binary instrumental variable to estimate the effect of outsourcing on productivity.

Keywords: *Endogenous selection, IV-estimation, inverse probability weighting, missing data, productivity, outsourcing, semiparametric estimation.*

1. INTRODUCTION

Missingness is a major problem in databases and survey-based data. While one well-known problem is recruiting a representative sample (unit non-response), a second major problem is incomplete answers (item non-response). This study focuses on item non-response, which arises when respondents to surveys prefer not to answer specific items or do not know the answer. Specifically, we focus on item non-response in the dependent variable. This problem particularly affects sensitive or specific information,

[1]We thank Irene Bertschek, Maximilian Kasy, Francois Laisney, Stephan Martin, Michael Ward for valuable suggestions. We are thankful to two anonymous referees for their comments. We also thank seminar participants at ZEW and Mines ParisTech for their questions and input.

which often are the outcome variables of interest at the heart of many economic studies. Examples are profits, turnovers, income, tax fraud, or the consumption of medications. A standard approach is assuming that data are *missing at random* (MAR) and then relying on listwise deletion of observations or on using imputation methods.[1] However, these practices may introduce a bias if the missingness is in fact selective, that is, when certain groups of observations are less likely to be reported/observed than others.

In this paper we propose a novel approach to correct for potential biases arising from missing data. Specifically, we study the estimation of averages of a selectively observed outcome in a cross-section context, when considering that certain characteristics are associated with higher frequencies of missing observations. The correction of the bias is based on an instrument that affects selection only through the partially observed outcome variable. In addition, we allow for endogeneity of regressors. We propose a simple two-step estimation procedure, and show that it is consistent and asymptotically normal.

We apply our estimator within a common class of econometric models, that is, production function estimation, and we study the effect of information technology (IT) outsourcing on productivity using a survey-based sample of German firms. Specifically, we aim to estimate the effect of IT outsourcing X on firm productivity Y^*, which is only partially observed. Commonly, in empirical studies the firms' sourcing decision is considered endogenous to the production process (e.g., Halpern et al. (2015); Görg et al. (2008); Amiti and Konings (2007)). However, focusing on endogeneity of the firms' sourcing activity is not sufficient in our application, since, in addition, the outcome Y^* is subject to selective missingness, as we illustrate in this paper. Indeed, firm productivity might directly influence the response behavior, for example, firms are less willing to report data after weak performance during the fiscal year. Consequently, we additionally have to correct for this selection error. To do so, we introduce an additional exclusion restriction on a control variable to account for selective item non-response. Specifically, we assume that this control variable does not contain any additional information on the missingness mechanism that is not already contained in the partially observed outcome Y^* and other controls. This exclusion restriction was recently considered by Ramalho and Smith (2013) and D'Haultfoeuille (2010). This assumption is suitable in situations in which selection is driven by the outcome Y^* itself. We argue that this is likely the case in many applications that rely on firm-level survey data.

Probably the most common approach to deal with missing observations is to assume *m*issing at random (MAR) (in the sense of Rubin (1976)), namely, that response depends only on observed covariates but not on the partially observed outcome variable. Unfortunately, the plausibility of this assumption may be questioned in many economic examples in which missing observations arise because of self-selection, or non-response, or because counterfactual variables are unobservable (for an analysis of sensitivity of MAR, see also Kline and Santos (2013)). In particular, when selection is driven by the underlying partially observed outcome itself, as we argue is likely with applications such as ours, existing empirical strategies that assume MAR are infeasible. When response is driven directly by the outcome, it might be also difficult to find instruments that determine selection but not the outcome (see Heckman (1974)). The idea of Heckman's instruments was also used in models with endogenous regressors (see Das et al. (2003)

[1]e.g. Heckman (1974), Rubin (1976, 1987).

for nonparametric estimation) but it is not suitable for our application as it requires instruments that explain the firm's response mechanism but not its productivity.

Contribution: In contrast to Heckman's instruments, we assume that IT outsourcing contains no information on the response mechanism that is not contained in potential productivity and other observed characteristics. Other existing methods that deal with selectively missing dependent variables by using a similar instrumental variable approach (see for instance Tang et al. (2003), Zhao and Shao (2015), or Miao et al. (2015)) cannot be applied, because of the endogeneity in the firms' sourcing activity. Using such instrumental variables when faced with endogeneity of both selection and covariates was studied by Breunig et al. (2018). While their approach leaves the functional form of the distribution of Y^* given covariates unrestricted, their solution requires continuity of the instrumental variables. However, the instrumental variables in the application of this paper are discrete and thus a different methodology is needed.

In this paper, we develop a new methodology that is particularly suited to our empirical application. In Section 2 of this paper we treat selective non-response of outcome and endogeneity of covariates in a partial linear model and establish identification given discrete instruments. Being able to use dummy variables as instruments for selective non-response in an IV estimation adds the last missing piece to render estimators that correct for selective non-response in the dependent variable fully functional. We propose a simple two-step estimation procedure: First, we propose a constrained nonparametric least squares estimator for the conditional selection probability of observing Y^*. Second, we enter this estimator in a generalized method of moments (GMM) estimator to arrive at the structural parameter. We implement the estimation procedure in Section 3 and estimate the effect of IT outsourcing on productivity. In this application, the instrumental variable is binary. We find that our estimation procedure performs well and effectively corrects for biases that would prevail when MAR is assumed. The method can be easily adopted to many applications using survey data. Section 4 provides a Monte Carlo simulation study and compares our estimator to an estimator based on the missing at random assumption.

Non random missingness is an important problem in the estimation of production functions. However, production function estimation has thus far focused on bias due to endogenous input choice and on endogeneity through panel attrition and unit non-response (firm exit) (cf. Olley and Pakes (1996), Levinsohn and Petrin (2003)), Melitz and Polanec (2015)). We highlight that imposing MAR on missing values in the dependent variable is an additional source of biased estimates and propose a correction that is compatible with IV estimation.

2. IDENTIFICATION AND ESTIMATION OF STRUCTURAL PARAMETERS

In this section, we provide assumptions under which the selection probability function and the conditional mean $\mathbf{E}[Y^*|X = \cdot]$ are identified. We further motivate our estimation procedure. For the sake of simplicity we first consider the situation in which the parametric part of our model consists of only a scalar endogenous regressor. Thereafter we discuss the situation in which the parametric part coincides with a vector.

2.1. Model

Our aim is to identify the causal impact of a binary, potentially endogenous variable X on a selectively observed outcome Y^*. We consider a partially linear model

$$Y^* = X\beta_0 + m(W_1) + U \tag{2.1}$$

for some unknown structural scalar parameter β_0 and unknown nonparametric function m. A realization of (Δ, X, W) with $W = (W_1', W_2)'$ is observed for each individual in the random sample. However, Y^* may suffer from selective non-response: a realization of the dependent variable Y^* is observed when $\Delta = 1$ and missing when $\Delta = 0$. We write $Y = \Delta Y^*$. Additionally, we model the exogenous covariates as W_1 and to deal with potential endogeneity in the explanatory variable of interest we allow for an instrument W_2 such that $\mathbf{E}[U|W] = 0$. Here, the instrument W_2 is binary.

Example 2.1. (IT Outsourcing and Productivity) *In our application, we use an augmented production function model to estimate the effect of IT outsourcing on productivity. In its stylized version, we consider the following model (we abstract from additional dummy variables and other controls):*

$$ln(Prod_i^*) = ITout_i\,\beta_0 + m\left(\ln(K_i), \ln(L_i)\right) + U_i.$$

In this empirical model $Prod_i^*$ denotes average labor productivity, which is only partially observed. We measure labor productivity by value added ($sales - costs\ of\ intermediates$) over labor. To avoid exceedingly complex notation, we simplify our application by considering capital K_i and labor L_i as exogenous control variables. Note that these variables are, in fact, strategic, and a full estimation should account for this complication.[2] *The parameter of interest is β_0, the coefficient that measures the effect of IT outsourcing $ITout_i$.*

Our possibly endogenous X variable is $ITout_i$, and the potential endogeneity is due to several reasons.[3] *We use a standard instrumental variable strategy, based on the excluded instrument W_2 to account for this. In our application, the instrument W_2 is a variable measuring whether a firm sought 'Y2K consulting' to avoid the 'millennium bug'.*[4] *See Section 3 for more details regarding our endogenous variable and the instrument in our application. The novelty of this paper is that we can model productivity $Prod_i^*$ to be plagued with selective item non-response, possibly because respondents avoid disclosing especially high (or low) value added. This is modeled by the response indicator Δ_i, which may depend on potential productivity itself. For instance, a company might be more likely to report data if its value added (and hence measured productivity) is high. Alternatively, when firms face complex situations and uncertainty, such as a negative productivity shock, they might have lower capacity or motivation to disclose sensitive financial information. Our strategy allows that the firm's response could be a function of value added. We*

[2] Appropriate corrections were suggested in the literature on production function estimation (e.g. Olley and Pakes (1996); Levinsohn and Petrin (2003); Wooldridge (2009); Petrin and Levinsohn (2012)). Yet, the purpose of our application is to illustrate how to implement our estimator that corrects for selective non-response in the dependent variable and its focus is the effect of IT outsourcing on productivity. Hence, while we deem it necessary to control for labour and capital, we avoid the additional expositional complexity that would arise from additionally correcting for endogeneity in labor and capital.

[3] For example, more productive firms could me more likely to use IT-outsourcing, or IT investment (and subsequent outsourcing) could depend on current profits. See Section 3 for more detail on this point.

[4] This bug threatened the IT systems of firms on January 1 2000, if they had relied on software that allocated only two 'year digits' when storing a date. See Section 3 for more detail on this excluded instrument.

show below that β_0 cannot be estimated consistently without accounting for the selectivity in the non-response for $Prod_i^*$.

2.2. *Identification*

In what follows, we show which general assumptions allow to identify β_0. Conditioning model (2.1) on the exogenous covariates W_1 yields

$$\mathbf{E}(Y^*|W_1) = \mathbf{E}(X|W_1)\beta_0 + m(W_1). \tag{2.2}$$

Multiplying equations (2.1) and (2.2) by the binary instrument W_2 and taking expectations leads to

$$\mathbf{E}(Y^* W_2) = \mathbf{E}(XW_2)\,\beta_0 + \mathbf{E}[m(W_1)W_2],$$
$$\mathbf{E}[\mathbf{E}(Y^*|W_1)W_2] = \mathbf{E}[\mathbf{E}(X|W_1)W_2]\,\beta_0 + \mathbf{E}[m(W_1)W_2].$$

Now taking the difference of both equations yields

$$\mathbf{E}[(Y^* - \mathbf{E}(Y^*|W_1))W_2] = \mathbf{E}[(X - \mathbf{E}(X|W_1))W_2]\,\beta_0. \tag{2.3}$$

The parameter β_0 is not identified if $\mathbf{E}(XW_2) = \mathbf{E}[\mathbf{E}(X|W_1)W_2]$. Identification of β_0 thus requires the instrument W_2 to contain information about X which is not captured by the exogenous covariates W_1. The next assumption formalizes this restriction.

Assumption 2.1. *It holds* $\mathbf{E}[(X - \mathbf{E}(X|W_1))W_2] \neq 0$.

Under Assumption 2.1 we can write the structural β_0 as

$$\beta_0 = \frac{\mathbf{E}[(Y^* - \mathbf{E}(Y^*|W_1))W_2]}{\mathbf{E}[(X - \mathbf{E}(X|W_1))W_2]}. \tag{2.4}$$

In the following, we provide sufficient conditions to ensure identification of $\mathbf{E}[(Y^* - \mathbf{E}(Y^*|W_1))W_2]$.

Assumption 2.2. (Exclusion Restriction on Selection) *It holds* $\Delta \perp\!\!\!\perp X \,|\, (Y^*, W)$.

Assumption 2.2 requires that the covariate X has no direct effect on the response given partially observed outcome variable Y^* and W. As such, the covariate X serves as an instrumental variable for the selective response to Y^*. Variable X is sometimes also called a shadow variable, see Miao et al. (2015). This assumption is well suited for our application but might also need to be modified to be appropriate for other particular applications. In fact, we may also assume that a subvector of W is independent of the response given the other potentially observed information. For instance, the instrument W_2 to account for endogeneity of X might also be used to account for selective non-response of Y^* via the exclusion restriction $\Delta \perp\!\!\!\perp W_2 \,|\, (Y^*, X, W_1)$. We thus can generalize the exclusion restriction to the condition $\Delta \perp\!\!\!\perp (X, W_2) \,|\, (Y^*, W_1, \widetilde{W})$ where $\widetilde{W}$ either coincides with W_2 (as in Assumption 2.2) or with X (as in the previous sentence).[5]

Example 2.2. (IT outsourcing and Productivity (cont'd)) *In our application Assumption 2.1, is satisfied if* $\mathbf{E}[(ITout - \mathbf{E}(ITout|Controls))Y2K] \neq 0$. *This means that the instrument*

[5]We are grateful to a referee for pointing out this generalization.

(which accounts for the endogeneity of IT outsourcing) contains information on IT outsourcing that is not captured by the other control variables. More importantly, Assumption 2.2 requires that $\Delta \perp\!\!\!\perp ITout \,|\, (Prod^*, L, K, Y2K)$, *that is, IT outsourcing* $ITout_i$ *contains no information on the response* Δ_i *that is not already contained in the potential productivity of firms,* $Prod_i^*$, *and observed control variables such as labor or capital.*

Example 2.3. (Relation to Triangular Model) *Assumption 2.2 can be justified in a triangular model as follows. Consider an equivalent formulation of model* (2.2) *as*

$$Y^* = \mathbf{E}(X|W_1)\beta_0 + m(W_1) + \varepsilon, \tag{2.5}$$

where $\varepsilon = Y^* - \mathbf{E}(Y^*|W_1)$. *In this case, the exclusion restriction on selection, i.e.,* $\Delta \perp\!\!\!\perp X \,|\, (Y^*, W)$, *is satisfied when additionally*

$$\Delta = \phi(Y^*, W, \eta),$$
$$\eta \perp\!\!\!\perp (X, \varepsilon) \,|\, W$$

for some unobservable η *and an unknown function* ϕ. *Similarly, different types of exclusion restrictions can be justified.*[6]

Assumption 2.2 implies $\mathbb{P}(\Delta = 1|Y^*, W, X) = \mathbb{P}(\Delta = 1|Y^*, W)$ and hence, by the law of iterated expectations, we obtain the following conditional mean restriction:

$$\mathbf{E}\left[\frac{\Delta}{\mathbb{P}(\Delta = 1|Y^*, W)}\Big| X, W\right] = 1. \tag{2.6}$$

The following assumption enables us to identify the conditional response probability $\mathbb{P}(\Delta = 1|Y^*, W)$ via the previous moment restriction. Let us denote $V^* \equiv (Y^*, W')'$, $V \equiv (Y, W')'$, and $Z \equiv (X, W')'$. Here, we denote $d_x = \dim(X)$ and $d_{w_1} = \dim(W_1)$.

Assumption 2.3. *(i) For all* v *in the support of* V^*, $\mathbb{P}(\Delta = 1|V^* = v) = G(v'\vartheta_0)$ *for some known strictly increasing function* $G : \mathbb{R} \to (0, 1)$ *and some parameter* $\vartheta_0 \in \mathbb{R}^{d_x + d_{w_1} + 1}$. *(ii) The parameter* ϑ_0 *is identified through* (2.6).

Assumption 2.3 (*i*) restricts the conditional probability of observing Y^* to be known up to a finite dimensional parameter. In particular, we model the selection probability in a single index framework. Typical examples are probit or logit models. Assumption 2.3 (*i*) also requires that the conditional probability of observing Y^* given (Y^*, X) is strictly positive. In particular, Assumption 2.3 can rule out certain types of selection, such as deterministic truncation models. Further, Assumption 2.3 (*ii*) ensures identification of the selection probability through equation (2.6) (see Theorem 2.1 of D'Haultfoeuille (2010) or Theorem 1 in Zhao and Shao (2015)). Note that the requirement of knowledge of G is not required if a completeness assumption of (Y^*, W) conditional on (X, W) is satisfied. Yet this would rule out binary instruments, as in our application, and hence, is not considered.

[6]For instance, the exclusion restriction $\Delta \perp\!\!\!\perp W_2 \,|\, (Y^*, X, W_1)$, mentioned above, is satisfied in model (2.5) when additionally

$$\Delta = \phi(Y^*, X, W_1, \eta)$$
$$\eta \perp\!\!\!\perp (W_2, \varepsilon) \,|\, (X, W_1).$$

EXAMPLE 2.4. (IT OUTSOURCING AND PRODUCTIVITY (CONT'D)) *In our application,* $\mathbb{P}(\Delta = 1|V^*)$ *denotes the probability that a company reports sales or costs of intermediates given potential productivity Prod* and other controls (that is* $\mathbb{P}(\Delta = 1|V^*) = \mathbb{P}(\Delta = 1|Prod^*, L, K, Y2K)$*). As we show below, identification of the function* $v \mapsto \mathbb{P}(\Delta = 1|V^* = v)$ *is key to identifying the structural parameter through inverse probability weighting but is also of interest on its own, because it provides evidence as to whether the MAR assumption is violated (see also Breunig (2017) for a formal test). In our application, we see that the conditional probability depends on the potential productivity realizations in a nonlinear fashion (see Section 3). We also show that reporting Prod* does depend on Prod* itself even if other important control variables are included, suggesting that the MAR assumption is violated.*

THEOREM 2.1. *Let Assumptions 2.1–2.3 be satisfied. Then, the structural parameter in model* (2.1) *is identified through*

$$\beta_0 = \frac{\mathbf{E}\left[\left(Y/G(V'\vartheta_0) - \mathbf{E}(Y/G(V'\vartheta_0)|W_1)\right)W_2\right]}{\mathbf{E}\left[(X - \mathbf{E}(X|W_1))W_2\right]}.$$

This concludes the identification argument, and we now move on to derive an appropriate estimator for our setting.

2.3. A Closed Form Estimator of the Structural Parameter

Our estimator of the structural parameter β_0 is based on the previous constructive identification results. We estimate the nuisance parameter ϑ and the nonparametric functions $g(w_1, \vartheta) = \mathbf{E}(Y/G(V'\vartheta)|W_1 = w_1)$ and $h(w_1) = \mathbf{E}(X|W_1 = w_1)$ in a first step. We replace ϑ_0 by a generalized method of moments (GMM) estimator $\widehat{\vartheta}_n$ based on an empirical analog of the conditional moment equation (2.6). The resulting estimator of β_0 then falls into the class of inverse probability estimators with estimated selection probability $G(V'\widehat{\vartheta}_n)$.

We propose the following new estimator of the structural parameter β_0 given by

$$\widehat{\beta}_n = \frac{\sum_{i=1}^n W_{2i}\left(Y_i/G(V_i'\widehat{\vartheta}_n) - \widehat{g}_n(W_{1i}, \widehat{\vartheta}_n)\right)}{\sum_{i=1}^n W_{2i}\left(X_i - \widehat{h}_n(W_{1i})\right)}. \tag{2.7}$$

where we replaced the nonparametric functions g and h by the series least squares estimators, as follows.

Let $L \geq 1$ denote the number of basis functions used to approximate these functions, where L increases with sample size n. We then introduce a vector of basis functions denoted by $p^L(w) = (p_1(w), \ldots, p_L(w))'$. Further, the matrix of basis vectors evaluated at the sample points of W_1 is denoted by $\mathbf{W}_n = (p^L(W_{11}), \ldots, p^L(W_{1n}))'$. We follow Breunig et al. (2018) and consider the following series least squares estimator with inverse probability weighting

$$\widehat{g}_n(w, \widehat{\vartheta}_n) \equiv p^L(w)'(\mathbf{W}_n'\mathbf{W}_n)^{-1}\sum_{i=1}^n \frac{Y_i}{G(V_i'\widehat{\vartheta}_n)}\, p^L(W_{1i}).$$

Moreover, we replace h by the series least squares estimator (see e.g., Newey (1997))

$$\widehat{h}_n(w) \equiv p^L(w)'(\mathbf{W}_n'\mathbf{W}_n)^{-1}\sum_{i=1}^{n} X_i\, p^L(W_{1i}).$$

The next result establishes the asymptotic distribution of the estimator $\widehat{\beta}_n$. We show its consistency with the identified structural parameter and asymptotic normality. In applications, such asymptotic distribution results can be useful in constructing approximate confidence intervals. The next theorem makes use of Assumption B.1 which is provided and discussed in the Appendix alongside the theorem's proof.

Theorem 2.2. *Let Assumptions 2.1–2.3 and B.1 be satisfied. Then we have*

$$\sqrt{n}\left(\widehat{\beta}_n - \beta_0\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

where σ^2 denotes the variance of the random variable

$$W_2\,\frac{Y/G(V'\vartheta_0) - g(W_1)}{\mathbf{E}[(X - \mathbf{E}(X|W_1))W_2]} - \Big(\frac{\Delta}{G(V'\vartheta_0)} - 1\Big)Z'\,A(A'A)^{-1}\,\mathbf{E}\left[\frac{W_2 - \mathbf{E}[W_2|W_1]}{\mathbf{E}[(X - \mathbf{E}(X|W_1))W_2]}YV\frac{G_\vartheta(V'\vartheta_0)}{G^2(V'\vartheta_0)}\right]$$

where $A = \mathbf{E}\left[ZV'G_\vartheta(V'\vartheta_0)/G^2(V'\vartheta_0)\right]$.

The asymptotic result of Theorem 2.2 remains valid if $\sqrt{n}\,(\widehat{\beta}_n - \beta_0)$ is normalized by the empirical analog of the variance σ^2, which follows by consistency of the variance estimator. This can be also used to construct pointwise confidence intervals for $\widehat{\beta}_n$. In our application, however, we rely on resampling methods.

2.4. Extension: Multivariate Control Variables

In the following extension, we lay out how our identification and estimation strategy carries over to models with multivariate endogenous regressors.[7] Again let (Δ, Y^*, X', W') be a jointly distributed random vector in which (Y^*, X, W) is a random vector that takes values in $\mathbb{R}^{1+d_x+d_w}$, and Δ is a random variable that takes values in $\{0, 1\}$. So, in contrast to the previous case, X is not scalar but a random vector and may also include exogenous covariates. As above, a realization of (Δ, X, W) is observed for each individual in the random sample while a realization of the dependent variable Y^* is observed when $\Delta = 1$ and missing when $\Delta = 0$ (again we let $Y = \Delta Y^*$). We consider a partially linear model

$$Y^* = X'\boldsymbol{\beta}_0 + m(W_1) + U \tag{2.8}$$

for some unknown parameter vector $\boldsymbol{\beta}_0$ and unknown nonparametric function m. In addition, to account for endogeneity of X, we assume that a multivariate instrument W_2 is available such that $\mathbf{E}[U|W] = 0$ where $W = (W_1', W_2')'$. As in the derivation of (2.3), conditioning model (2.8) on the exogenous covariates W_1 and/or the instruments W_2 yields

$$\mathbf{E}[Y^* - \mathbf{E}(Y^*|W_1)|W_2] = \mathbf{E}[X - \mathbf{E}(X|W_1)|W_2]'\,\boldsymbol{\beta}_0 \tag{2.9}$$

[7] Note that this extension is beyond the scope of our specific application.

The parameter vector $\boldsymbol{\beta}_0$ is not identified if $\mathbf{E}(X|W_2) = \mathbf{E}(\mathbf{E}(X|W_1)|W_2)$. Intuitively, this means that W_1 has no additional information for explaining variation in X that is already available from W_2. Put differently, identification of β_0 requires the instrument W_2 to contain additional information in explaining variations of X that is not included in the exogenous covariates W_1. The next assumption formalizes this restriction.

Assumption 2.4. *The matrix* $\mathbf{E}\Big[\,\mathbf{E}[X - \mathbf{E}(X|W_1)|W_2]\,\mathbf{E}[X - \mathbf{E}(X|W_1)|W_2]'\Big]$ *is invertible.*

To check Assumption 2.4 we refer to the rank tests considered in the literature, see, for instance, Kleibergen and Paap (2006) (given the additional difficulty of estimating $\mathbf{E}(X|W_1)$). In the following, let us introduce the vector valued function $\mathbf{h}(w_1) = \mathbf{E}(X|W_1 = w_1)$. Assumption 2.4 ensures that the $\boldsymbol{\beta}_0$ is identified through equation (2.3) (given that the left-hand side is identified), and we can write

$$\boldsymbol{\beta}_0 = \Big(\mathbf{E}\Big[\,\mathbf{E}[X - \mathbf{h}(W_1)|W_2]\,\mathbf{E}[X - \mathbf{h}(W_1)|W_2]'\Big]\Big)^{-1}\mathbf{E}\Big[(Y^* - \mathbf{E}(Y^*|W_1))\,\mathbf{E}[X - \mathbf{h}(W_1)|W_2]\Big]$$

In the following, we provide sufficient conditions to ensure identification of $\mathbf{E}[(Y^* - \mathbf{E}(Y^*|W_1))W_2]$. Since, by Assumption 2.3, it holds

$$\mathbf{E}(Y^*|W) = \mathbf{E}\left[\frac{Y}{G(V'\vartheta_0)}\Big|W\right] \tag{2.10}$$

where the right-hand side is identified, we obtain the following identification result of the multivariate structural parameter $\boldsymbol{\beta}_0$.

Proposition 2.1. *Let Assumptions 2.1–2.4 hold true. Then, in model* (2.8), *the parameter* $\boldsymbol{\beta}_0$ *is identified through*

$$\begin{aligned}\boldsymbol{\beta}_0 = &\Big(\mathbf{E}\Big[\,\mathbf{E}[X - \mathbf{h}(W_1)|W_2]\,\mathbf{E}[X - \mathbf{h}(W_1)|W_2]'\Big]\Big)^{-1}\\ &\times \mathbf{E}\Big[\Big(Y/G(V'\vartheta_0) - \mathbf{E}(Y/G(V'\vartheta_0)|W_1)\Big)\,\mathbf{E}[X - \mathbf{h}(W_1)|W_2]\Big].\end{aligned}$$

In order to estimate the parameter vector of interest $\boldsymbol{\beta}$, we need to estimate the nuisance parameter ϑ and the functions $g(w_1,\vartheta) = \mathbf{E}(Y/G(V'\vartheta)|W_1 = w_1)$ as above and the vector valued function $\mathbf{h}(w_1) = \mathbf{E}(X|W_1 = w_1)$ in a first step. We replace ϑ_0 with a GMM estimator $\widehat{\vartheta}_n$ based on an empirical analog of the conditional moment equation (2.6). Further, we replace the function $\mathbf{h}$ with the series least squares estimator

$$\widehat{\mathbf{h}}_n(w) = p^L(w)'(\mathbf{W}_n'\mathbf{W}_n)^{-1}\sum_{i=1}^n X_i p^L(W_{1i})$$

In a second step, we propose the estimator of $\boldsymbol{\beta}$ given by

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_n = &\left(\sum_{i=1}^n \Big(X_i - \widehat{\mathbf{h}}_n(W_{1i})\Big)\Big[W_{2i}(\mathbf{W}_{2n}'\mathbf{W}_{2n})^{-1}\sum_{i'=1}^n W_{2i'}\Big(X_{i'} - \widehat{\mathbf{h}}_n(W_{1i'})\Big)\Big]'\right)^{-1}\\ &\times\sum_{i=1}^n \Big(Y_i/G(V_i'\widehat{\vartheta}_n) - \widehat{g}_n(W_{1i})\Big)W_{2i}(\mathbf{W}_{2n}'\mathbf{W}_{2n})^{-1}\sum_{i'=1}^n W_{2i'}\Big(X_{i'} - \widehat{\mathbf{h}}_n(W_{1i'})\Big)\end{aligned}$$

where we make use of the notation $\mathbf{W}_{2n} = (W_{21},\ldots,W_{2n})'$. This concludes our extension for multivariate control variables, and we now turn to applying our estimator to a production function setting using survey data.

3. APPLICATION: THE IMPACT OF IT OUTSOURCING ON FIRM SUCCESS

3.1. Setting and Motivating Question

We now apply our estimation procedure developed in Section 2 to study the effects of IT outsourcing on firm performance using firm-level micro data. We follow the empirical literature on services outsourcing (see below) and study IT outsourcing using an extended production function framework.

Selective item non-response in firm-level data In settings like ours, high item non-response rates in particular variables complicate identification of the model parameters. Items that are plagued by a considerable share of non-response are typically monetary values, such as sales, or costs of intermediate inputs, which are required to construct key variables of interest. This problem has been documented for many business surveys that are fundamental to economic research. An important example is the US Census Bureau's Census of Manufacturers (CM), which is the main data source for much of the research on US plant-level productivity.[8] White et al. (2012) document shares of imputed values in 2007 for the items' total value of shipments, cost of electricity, and cost of material inputs of 27%, 37%, and 42%, respectively. The situation is similar for the establishment panel of the Institute for Employment Research of the Federal Labor Service in Germany (IAB), which is a cornerstone database for firm-level research in Germany. Some of the highest rates of non-response in the 2007 wave of the survey arise for key variables such as payroll (14.4%), intermediate inputs' share of revenue (17.4%), and last year's annual revenue (18.6%) (Drechsler (2010)). These high rates of item non-response in key variables highlight the scope of the problems that item non-response might cause if it depended on the undisclosed variable's value.

In addition to firm-level survey data, the problem of missing values in items referring to monetary values is also well documented for individual and household surveys.[9] Particulary in the context of firms, non-response might result from the lack of the right information for the individual respondent being surveyed. In addition, non-response, specifically for monetary values, is frequently related to the perceived sensitivity of the information (Kennickell (1998); Drechsler (2010)). According to Tomaskovic-Devey et al. (1994), managers often have doubts about the confidentiality of surveys and refuse to disclose confidential financial information requested, which is the most important reason for firms' non-response.

While evidence on firm-level non-response is scarce, existing studies focusing on unit non-response suggest that MAR can be violated within business surveys, and thus provide additional motivation for our methodology to remedy potential bias from selective item non-response. In particular, our focus on selective non-response is in line with findings of Earp et al. (2014), who examine unit non-response in dependence on

[8] For instance, Olley and Pakes (1996); Foster et al. (2008); Black and Lynch (2001).

[9] See, e.g., Frick and Grabka (2010) for its discussion of non-response issues in earnings and wealth variables in the German Socio-Economic Panel (GSOEP), the British Household Panel Survey (BHPS), and the Household, Income and Labor Dynamics in Australia (HILDA) survey. Kennickell (1998) examines the issue in the context of the Survey of Consumer Finances (SCF).

firm characteristics. Applying a regression tree framework to survey data for companies in the American agricultural sector, the authors show that total sales, as well as total value of products sold are the strongest predictors for unit non-response. What is more, similar to the relation between productivity and selective missingness uncovered by our methodology, the authors show that unit non-response is negatively dependent on total sales, even after controlling for other characteristics such as firm size. In another related study investigating a sample of Italian firms, Borgoni et al. (2012) find that firms with extreme values in sales in both tails of the distribution exhibit a lower propensity to participate at all in the survey after controlling for industry and firm size.

In dealing with item non-response, applied empirical research based on firm-level survey data commonly rests on assuming MAR and pursues listwise deletion or is based on imputed data. We argue that assuming MAR likely results in biased estimates in applications such as ours for two reasons. The probability of response in business surveys can often be related to factors such as unit size or industry. Small firms may not keep track of requested items, because of lower reporting obligations (see, e.g., Thompson and Washington (2013)). But also large and more complex organizations might respond with lower probability. The probability to interview the right individual with access to the requested information and their authority to report might be lower. What is more, apart from the relevance of possibly unobserved firm characteristics, we stress that, in many cases, item non-response is likely to be heavily driven by the underlying latent variables themselves. For example, at firms that experienced negative shocks and generated low sales over the fiscal year, the respondents might wish to keep poor performance confidential, given the often stated perceived sensitivity of financial information. Overall, when firms have to deal with complex situations and uncertainty, as is the case after a negative productivity shock, they might have lower capacity or motivation to disclose sensitive financial information (see Tomaskovic-Devey et al. (1994)). In such cases, the MAR assumption will be violated, and commonly used strategies in applied empirical research (listwise deletion and imputation) will yield biased estimates.

Related Literature The theoretical literature on outsourcing dates back to the seminal work by Coase (1937) and his theory of the firm. Traditionally, this literature focuses on transaction costs and incomplete contracts (Williamson (1989, 1981, 1979); Grossman and Hart (1986)) to explain vertical integration. More recent literature focuses on the rise in services outsourcing in response to a rapid expansion of the business services sector and trade (see, e.g., Abraham and Taylor (1996); Feenstra (1998); Grossman and Helpman (2005)). While theory motivates international outsourcing primarily by differentials in factor prices, it explains domestic services outsourcing by scale economies of specialized input providers. Outsourcing might also help to even out the workload of the workforce when demand is volatile (Abraham and Taylor (1996)).

IT outsourcing has been a key dimension of business services outsourcing, at least since Eastman Kodak handed its entire data and microcomputer operations to an IBM-led consortium (Loh and Venkatraman (1992)). This is not surprising, given the great importance of information technology for productivity, which has been widely documented both for the wider economy (Brynjolfsson and Hitt (2003a)), and particularly for information intensive sectors such as the health sector (Lee et al. (2013)).[10] The importance of IT outsourcing is reflected in its steady growth over the past few decades (ZEW (2010); Han et al. (2011)). Outsourcing IT services can be an attractive way to leverage cost advantages, and it can also facilitate the restructuring of production such that the remaining workers become more efficient (Amiti and Wei (2009)). Moreover, drawing on more specialized providers can increase the quality of IT services and thus improve input quality (Lacity et al. (2009)).

Against this background, we investigate whether IT outsourcing increases labor productivity at German manufacturing and services firms. This question has important policy implications for both investment in more powerful IT infrastructure and labor policy.

3.2. Empirical strategy

In order to investigate the effect of IT outsourcing on firm-level average labor productivity, we estimate a production function augmented by firms' IT outsourcing activities. In particular, we model labor productivity at firm i, $Prod_i^{*}*$, as a function of capital, K_i, and labor inputs, L_i. IT outsourcing, $ITout_i$, is a binary variable indicating whether the firm subcontracted IT services and enters our production function as a shift parameter (alongside other controls W_1):

$$ln(Prod_i^*) = m(ln(K_i), ln(L_i)) + \beta_0 ITout_i + W_{1,i}'\beta_1 + u_i. \tag{3.11}$$

In line with Equation (2.1), we allow our production function to be flexible with respect to capital and labor inputs. Assuming $m(.)$ to be linear in K and L gives the empirical production function based on an augmented Cobb-Douglas production function as a special case.[11] We estimate both the partially linear and the linear Cobb-Douglas model (the latter using two-stage least squares). In all estimations we allow for endogeneity of IT outsourcing.

Estimation of production functions such as equation (3.11) is often complicated by considerable item non-response in the measures of output or value added, which is typically constructed from data on firms' financial performance. As we expect item non-response in measures for labor productivity to be driven by the underlying value, we expect that MAR is commonly violated in comparable empirical applications using survey data. We therefore resort to our estimation strategy developed in Section 2.

Exclusion Restriction 1: For this approach, we impose an exclusion restriction that relies on independence between firms' outsourcing status ($ITout_i$) and Δ_i conditional on ($Prod_i^*, K_i, L_i, Y2K_i$).

Unlike the exclusion restrictions in a standard IV, the exclusion restriction that allows to handle selective item non-response is in principle testable as shown by D'Haultfoeuille (2010). However, finite sample power of such a test might be low in practice, so that it

[10] Earlier studies have found similarly important productivity contributions in the health sector (Menon et al. (2000)), or in retailing (Schreyer and Pilat (2001); Reardon et al. (1997)).

[11] We start with a standard Cobb-Douglas production function $Y_i^* = A_i K_i^{\alpha_K} L_i^{\alpha_L} ITout_i^{\beta_0}$ with output Y_i^* being a function of capital K_i, labor inputs L_i, and a Hicks-neutral efficiency term A_i. The binary variable $ITout_i$, indicates use of IT outsourcing and enters the production function as a shift parameter. Dividing by L_i, taking logs on both sides and adding an i.i.d. error term u_i gives the linear version of empirical model.

seems warranted to inspect the assumption's credibility in greater detail. For exclusion restriction 1 to be violated a firm's outsourcing status would have to carry additional information about the interview partners response behavior beyond our control variables and our performance measure. This would be expected if an unobserved factor influenced both a responders "non-response" on items which are used to derive the dependent variable in the value-added production function and the probability of outsourcing.

Several factors might influence a responders item non-response on productivity. First, it depends whether the survey is answered by the owner or an employee. The owner might be more knowledgeable, whereas an employee might simply not have the requested information available. Even if they have the information, employees might be uncertain, whether management considers the information on financial measures confidential, and especially so, if the firm is experiencing a negative shock. However, for such patterns to violate exclusion restriction 1, the identity of the respondent would have to systematically correlate with outsourcing. We can see such a correlation through firm size, when the complexity of the organization affects the chance of interviewing the individual with apt knowledge and authority to provide information and firm size correlates with outsourcing of business processes. However, we control for firm size, and thus account for this channel.

Other factors that might influence item non-response could be sector specific behavior or practices regarding how openly firms talk about their profits. This would be a threat to our exclusion restriction 1, but we account for sectors in our estimation. We could also think that regional conventions that affect how openly people talk about money might influence the response behavior. However, first, such conventions could only threaten exclusion restriction 1 if it is systematically correlated with factors that influence outsourcing, such as attitudes towards foreigners, and, second, we control for the regions to some extent, and we have not observed such region-specific response behavior. Another concern, could be if IT outsourcing were correlated with accounting outsourcing. If accounting is outsourced, one may expect that people within the firm may be less able to answer questions on total sales or intermediary inputs. In combination with a correlation between the two types of outsourcing, this would lead to a violation of the exclusion restriction. For Germany in 2004, programming and IT-outsourcing were much more common than outsourcing of accounting or other business processes, and these two types of outsourcing were relatively orthogonal decisions.[12] As a result, only a low share of firms in Germany outsourced their accounting (Ohnemus (2009)).

In sum, we consider it unlikely that the firms outsourcing status carries additional information about the interview partners response behavior beyond our control variables and our performance measure. We thus expect our exclusion restriction 1 to hold.[13]

Exclusion Restriction 2: In addition to accounting for selective item non-response in $Prod_i^*$, we allow for endogeneity of the outsourcing decision in Equation (3.11) via a standard IV approach. For that we use *Y2K consulting* as excluded instrument, which

[12]For data from 2007, the observed correlation between the outsourcing of IT-services and of accounting was weakly negative (-0.0867, at the 10%-level) across the 3957 firms.

[13]We note, however, that very few exclusion restrictions are truly without limitations, and hence small doubts might ultimately remain. A viable alternative to exclusion restriction 1 would be using $\Delta \perp W_2 | (Y^*, W_1, X)$. If W_2 is truly random, it should not affect directly non response on Y^*, and would hence be a valid alternative to the route chosen in this paper. We are grateful to an anonymous referee for pointing to this issue.

measures whether a firm resorted on external consultancy for the year 2000 problem (also known as the Y2K problem, or the millennium bug; cf. Ohnemus (2007)). The year 2000 problem was due to "short sighted" early computer programming, which stored only the last two digits of a year. This practice would have caused some date-related processes to operate incorrectly from January 1, 2000, onwards. The wider public was essentially unaware of this issue until 1997. Virtually all firms were equally confronted with the year 2000 problem, but the extent of the required consulting services depended on how seriously the Y2K problem affected any given firms workflow. This instrument is relevant, because suffering from the Y2K problem may increase the likelihood of using IT outsourcing and relying on external services to solve the problem. After a firm gains experience in using external help to solve IT problems, management might be more inclined to outsource other IT activities as well.

Exclusion restriction 2 holds if the year 2000 problems are unrelated to a firms productivity in 2004. This assumption is plausible, because (a) the Y2K problem was unexpected until 1997, and (b) the variation in the size of the problem between firms depended in part on the support by the technology's provider, which was outside the scope of the firm, and thus introduced a random element into the likelihood with which firms hired consultants. In particular, ICT manufacturers reacted by compiling warning lists about possible bugs in their products and provided patches and updates. The severity of the Y2K problem for a particular firm thus depended on the reliability of the respective ICT supplier and the information and support they provided. Exclusion restriction 2 would be violated if management focused on an operating systems use of two or four digits to store years at the time the purchase decision was made. However, whether years are stored with two or four digits is a deep feature of programming, which did not receive broad media attention until the end of the 1997, and the problem was totally unexpected. In fact, in the US, for instance, hardly any business with less than 2000 employees had taken any measures towards this problem until 1997 (U.S. Department of Commerce (1999)). We consider such a managerial focus was unlikely.

Firm-size is another factor that might drive both Y2K-related outsourcing and productivity. In larger firms the IT department, rather than management, might take the decision to use Y2K consulting. This may simply imply that the Y2K instrument is even stronger for large firms, but even if this could introduce a confounding element, we control for firm-size. A firm's age might also play a role here, since younger firms might be less affected, if they use newer computer systems. On the other hand, the real driver is the software that the company used, and several deeper lying heritage systems could still be affected. It is also possible that affected firms engaged in both consulting and an inventory shift, so that they might have newer, better quality IT afterwards. While this would be an omitted variable to worry about, we control for general ICT-intensity by the "share of employees working with PC."

Another concern is that the firms' reaction to the Y2K bug might be dependent on the industry they are operating in. In this way, some industries are more dependent on electronic transactions and more sensitive to data security issues than others. Moreover, firms in well regulated or competitive industries with little scope to pass on potential costs of failures to customers might have stronger incentives to take extensive precautionary actions than firms in more concentrated industries where there is greater scope to pass on costs (U.S. Department of Commerce (1999)). As we account for industries in the analysis, this channel is unlikely to violate exclusion restriction 2.

Finally, exclusion restriction 2 could also be violated if hiring y2k-consulting were to be correlated with general managerial characteristics and, in particular, with the firm's general experience with outsourcing. We can address this concern in a robustness check and show the results in Appendix C. In this Appendix, we leverage the fact that we observe the firms' use of general outsourcing on a subsample of 1,246 firms that were surveyed in the previous round of the survey. We use these data to investigate to which extent general experience of the management with outsourcing could interfere with our estimation strategy: First, general outsourcing and y2k-consulting are not significantly correlated, conditional on the other control variables. This highlights the randomness in whether or not firms had to rely on external consultancy for the specific problems caused by the millenium bug.[14] Second, if we include general outsourcing as control our main results are qualitatively confirmed, with the exception that the bias of not correcting is seen even more clearly in the smaller sample. Specifically, we find insignificant IV-estimates for IT outsourcing if we fail to apply the correction we propose in this paper. This is also true if we replicate our main specification on the sample of firms for which we observe general outsourcing. Taken together, this additional evidence suggests that y2k-consulting is not driven by the same underlying factors as general outsourcing.

3.3. *Implementation Details*

We implement our semiparametric estimator, which we derived in the previous section. In the first step of our estimation procedure, we estimate the selection probability function, which is used in the second step to weight the observations in the actual production function estimation. The second-step estimation applies these weights, but otherwise uses only those observations for which the dependent variable *Prod** is observed, i.e., when $\Delta = 1$.

For the first step of the estimation, we need to introduce a link function G for a parametric model of the response mechanism Δ. The function G chosen coincides with the cumulative standard normal distribution Φ. Further, because of the estimation of the conditional probability $\mathbb{P}(\Delta = 1|V^* = v) = \Phi(v'\vartheta_0)$ we face a nonlinear optimization problem. To do so, we adopt the following choice of the starting value.

1 Estimate the parameter ϑ_s under missing completely at random (MCAR), i.e., the first entry of the parameter vector is the empirical analog of $\Phi^{-1}\big(\mathbb{P}(\Delta = 1)\big)$ and all other parameters are set to zero. In our application, we chose the first entry of ϑ_s somewhat smaller to ensure convergence of our optimization routine.

2 Linearize the estimation problem through a first-order Taylor approximation around ϑ_s, i.e.,

$$\mathbf{E}\left[\Big(\frac{\Delta}{\Phi(V'\vartheta)} - 1\Big)Z\right] \approx \mathbf{E}\left[\Big(\frac{\Delta}{\Phi(V'\vartheta_s)} - 1\Big)Z\right] - \mathbf{E}\left[\frac{\Delta V'Z\,\varphi(V'\vartheta_s)}{\Phi^2(V'\vartheta_s)}\right](\vartheta - \vartheta_s),$$

where φ is the standard normal probability density function. The norm of the linearization is minimized when ϑ coincides with

$$\vartheta^* \approx \vartheta_s + \mathbf{E}\left[\frac{\Delta V'Z\varphi(V'\vartheta_s)}{\Phi^2(V'\vartheta_s)}\right]^{-1}\mathbf{E}\left[\Big(\frac{\Delta}{\Phi(V'\vartheta_s)} - 1\Big)Z\right],$$

[14] When regressing y2k-consulting on general outsourcing and the full set of our control variables (reported in the Appendix), the estimated coefficient for general outsourcing is -0.004 with a standard error of 0.032.

where we used that $\dim(V) = \dim(Z)$, which is satisfied in our application.

Moreover, our semiparametric estimation approach relies on the choice of smoothing parameter L used in our estimator $\widehat{g}_n$ and $\widehat{h}_n$ (see Section 2.3), which is implemented via cross validation.[15] Note that cross validation does not lead to a theoretically accurate undersmoothing, which is required for the asymptotic distribution result. However, this method for the choice of L has a tendency of undersmoothing in finite samples and is, hence, broadly applied for the implementation of inference results. For the estimation of the finite sample variance of our estimator we use the bootstrap, i.e., the empirical variance of the bootstrap estimators derived from each resampling step. An alternative procedure relies on multiplier bootstrap as considered in Breunig et al. (2018).

3.4. Data Description and Summary Statistics

We use data from a firm survey conducted via computer-aided telephone interviews by the Centre for European Economic Research (ZEW). The survey has a special focus on the diffusion and the use of information and communication technologies (ICT) at German companies. For our application, we use the 2004 wave of the data, which contain information on firms' IT outsourcing activities. The sample is drawn using a stratified sampling design, with stratification cells being defined by size class of the firm, industry affiliation, and two regions (East/West Germany).[16] In order to use *Y2K consulting* as excluded instrument in a standard IV approach, we effectively consider only firms that already exist before 2000, and are thus older than four years in 2004.

We follow the usual approach taken in the literature and measure firms *average labor productivity* by total sales minus costs of intermediate inputs (in euros) per employee, $Prod = (sales - costs\ of\ intermediates)/L$. Missing values in average labor productivity stem from considerable item non-response to the two survey questions on *total sales*, as well as on the share of sales attributed to intermediate inputs and external costs (*cost share*) during the fiscal year.

The survey questionnaire covered the entire range of IT services that companies might need to operate their business, asking further whether the firms had outsourced each specific activity to an external service provider in whole or in part. We restrict the analysis to services that are required at every firm using computer technology in its business operations, namely, the (i) installation of hardware and software, (ii) computer system maintenance, and (iii) user assistance and support.[17] The constructed dummy variable for *IT outsourcing* used in our estimation takes the value of 1 if a firm outsources at least one of those three basic IT services completely and 0 otherwise.

As is the common case in respective firm-level survey data, no information is available to directly measure the physical capital stock of the firms. We therefore assume investment to be proportional to the capital stock and use *gross investment* figures as

[15] Specifically, we use quadratic B-splines and cross validation yields the choice of two knots for either $\widehat{g}_n$ and $\widehat{h}_n$, hence $L = 4$.

[16] As a sampling frame, the survey uses the data pool of "Verband der Vereine Creditreform" (CREDITREFORM), a credit rating agency, which provides the largest database on firms available in Germany. For more information on the data and data access see Bertschek et al. (2017).

[17] We thereby disregard more sophisticated IT services, which most of the firms in our sample do not need, such as software programming.

an empirical proxy for capital K (see, e.g., Raymond et al. (2015); Bertschek and Kaiser (2004)). We measure *labor L* in full-time equivalent terms, assuming that a part-time employee represents half of a full-time employee. The instrumental variable chosen for *Y2K consulting* is a dummy variable that equals 1 if a firm resorted on external consultancy for the year 2000 problem (0 otherwise). We additionally control the firms' overall IT intensity. Thus, we include the share of employees working predominantly with personal computers in the model (*pcwork*). This measure is a common proxy for 'general purpose' IT and has been widely used in the IT productivity literature (e.g., Bloom et al. (2012); Bresnahan et al. (2002)). Moreover, we include 13 industry dummies constructed from two-digit standard industry codes (NACE)[18] and a dummy indicating whether the firm is located in Eastern Germany.

The raw data for this paper consist of 3,801 observations.[19] In most items the share of missing values is well below 5%. In addition to our dependent variable, reported investments stands out, with about 27% missing observations in the raw data (see Table 2). For variables with modest missingness (rates below 5%), we regard the assumption of MCAR and applying listwise deletion as innocent. Additionally correcting for item non-response in an independent variable is possible, but would add considerable weight to our exposition. Hence, for this application, we assume MCAR for all independent variables and perform a complete case analysis in (X, W).[20] This simplification allows us to focus on the potential bias due to item non-response in the dependent variable $Prod^*$ and how it can be corrected. We stress, however, that MCAR in investments is a strong assumption. A full estimation should correct for potential non-selective missingness in this variable.

Table 3 reports summary statistics of the resulting estimation sample, consisting of 2,631 observations for which (X, W) is fully observed. In our estimation sample $Prod^*$ is missing for 535 observations. As the number of employees is fully observed, the missing observations in $Prod^*$ stem from item non-response to survey questions on sales and the cost share of intermediate inputs. Overall, information on sales is missing for 276 observations and information on the cost share is missing for 349 observations. Information on both items is missing for 90 observations. Consequently, the incidence of missingness in the dependent variable in our estimation sample totals about 20%. For reasons outlined in the above sections, we expect the missingness in $Prod^*$ not to be random and the resulting bias to be far from negligible, given the considerable share of item non-response.

3.5. Results

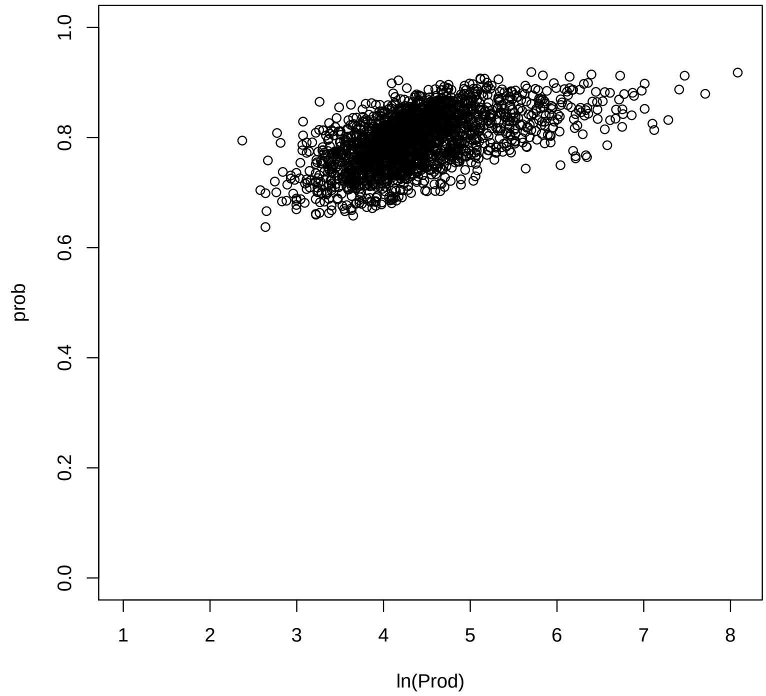
This section outlines the application of our estimation procedure. We also evaluate our estimator against the assumption of MAR. As a benchmark, we use listwise deletion

[18] See Table 1 for the industry distribution of the estimation sample.

[19] The complete survey data include 4,252 observations. We drop 369 observations from the sector 'electronic processing and telecommunication', because firms providing IT services to other companies typically belong to this sector and cannot be meaningfully included in the analysis. We further removed 82 observations with illogical values in input and output measures to arrive at a dataset of 3,801 valid observations.

[20] We could easily mirror datasets, such as the Census of Manufacturers (CM), and impute missing values in the investment variable White et al. (2012), but this would still require assuming MAR, or alternatively further distributional assumptions about the data or the response mechanism (see e.g. Paiva and Reiter (2014) for a Bayesian approach).

Figure 1. Conditional Probability of Item Non-response and Observed Outcomes.



NOTES: Figure 1 illustrates the first-step estimation using all 2,631 observations. It displays the estimator of the function $v \mapsto \mathbb{P}(\Delta = 1|V^* = v)$ evaluated at the realizations $(Prod_i, L_i, K_i, Y2K_i)$ when productivity is observed, i.e., $\Delta_i = 1$. The figure plots the estimated conditional probability against the observed realizations of $Prod^*$.

as well as multiple imputation (Rubin (1978)), which in practice is the most commonly used strategy in dealing with item non-response in practice.

The first step of our estimation procedure accounts for non-random missingness of the dependent variable. In this step, we estimate the selection probability function, which makes it the critical step of our method. This estimation step involves the outcome measure as well as the indicator for IT outsourcing, $IToout$, which we use as our instrument for selection to model the response mechanism Δ. In addition, we include capital K, labor L, and $Y2K$, the indicator for Y2K-consulting, into the first-step estimation. In the second step, we then use the parameter estimates $\widehat{\vartheta}_n$ to compute the weighing factor of each observation. The weighing factors are then used to weight each observation in the actual production function estimation, which uses only observations for which the dependent variable $Prod^*$ is observed.

Based on our first-step estimation results, Figure 1 shows how the probability of item non-response is related to observed realizations of the latent dependent variable $Prod^*$. From that Figure we suspect - without providing a formal test - that the non-response is not random, but becomes less likely for larger values in $Prod$. Therefore, Figure 1 clearly suggests that MAR could be violated. The first-step estimation results thus highlight the need to account for the nonignorable non-response in $Prod^*$ in the second step of our estimation procedure, which is in line with existing findings on selective unit non-response (Borgoni et al. (2012); Earp et al. (2014)). In particular, as we observe a positive relationship between the probability of observing $Prod^*$ and the underlying value itself, estimation strategies relying on MAR are based on response probabilities that are too large. Thus, using MAR techniques, we would likely underestimate the true population parameter β_0.

Table 5 shows our main results from the second step of our estimation procedure. All specifications estimate variants of the model discussed in Section 2. We show two groups of three columns. Columns 1–3 show the linearized version of the production function model in Equation (3.11). We estimate the model by two-stage least squares.[21] Columns 4–6 correspond to the more flexible, and preferred partially linear model. This specification does not impose linearity on $m(\cdot)$. Columns 1 and 4 estimate the model, assuming that the dependent variable was MCAR and deleting the entire observation from the estimation (listwise deletion). Columns 2 and 5 use imputation techniques which assume the variable was missing at random (MAR) to keep the observation in the dataset. Columns 3 and 6 apply the correction developed in Section 2. Hence, column 6 shows our preferred estimator, which combines the correction for selective missingness with the inclusion of K and L in the non-parametric component of the model.

In all specifications, we control for labor and capital, and include 13 sector dummies, an indicator for a firm's location in Eastern Germany, as well as the measure for the firm's IT intensity in the second-stage estimation of the production function. We report bootstrap standard errors obtained using 500 repetitions. As suggested by the first-step results in Figure 1, we expect that assuming MCAR (columns 1 and 4) leads to much smaller coefficient estimates. In columns 2 and 5, we attempt to correct for the bias of MCAR by using multiple imputation. We impute $ln(Prod^*)$ using all variables available in our estimation sample as predictors.[22]

In columns 3 and 6, we use the full estimation procedure that we propose in this paper. While, in column 3, we apply the correction within the linear specification of the production model, column 6 shows the results for the partially linear model, which underlies the discussion in Section 2. Our correction for the selective missingness leads to considerably larger coefficient estimates in both the linear and the partially linear model. Given the positive relation between the response and the underlying value of the outcome, and the positive relationship of the outcome and outsourcing, we would expect estimation based on listwise deletion (MCAR) or imputation (MAR) to underestimate the effect of outsourcing. While the correction leads to somewhat wider confidence bands, the standard errors are reasonably small to guarantee meaningful inference.

Regarding the interpretation of our result, we find positive and economically meaningful productivity returns to IT outsourcing (in all specifications). However, these

[21] See Table 4 for the estimation results of the first stage with and without the excluded instrument.

[22] The imputation was conducted using the *R* package *mi* (`https://cran.r-project.org/web/packages/mi/`). We apply an additive linear model to impute $ln(Prod^*)$. Due to the skewness of some continuous variables, we log-transform the dependent variable, as well as the labor and capital variable in the imputation model. All other variables enter the model in levels. We generate $m = 5$ datasets and combine the individual estimation results according to Rubin (1978). Note that we could also impute the underlying items sales and costs, rather than imputing the dependent variable $ln(Prod^*)$ directly. Separately imputing sales and cost would allow us to exploit additional information in the data in cases when only one of the underlying items is missing. However, this procedure would provide one of the methods with more information than the others. Thus, for the sake of comparability between listwise deletion, imputation, our estimator and the respective underlying assumptions we refrain from imputing sales and cost separately. For the same reason we impute $ln(Prod^*)$ using our estimation data, rather than imputing all missing variables in the raw data. However, when MAR is violated, the imputation insufficiently corrects for the bias. In fact, we find the estimation results based on multiple imputations of $Prod^*$ to be close to the results obtained by listwise deletion.

positive returns are underestimated under the MAR assumption and when employing standard methodologies, such as listwise deletion (MCAR) and multiple imputation. These results need to be interpreted carefully though, because we maintain the restrictive assumption that both labor and capital are exogenous inputs to the production function. This simplification is only useful for the purpose of the present application, which is providing a valid illustration of how the proposed estimator can be implemented in practice.[23] While our coefficients are not completely unreasonable, they do deviate somewhat from commonly reported estimates. Our coefficients are in line with constant returns to scale, but our preferred estimates of the elasticity of productivity with respect to capital is on the lower end of the spectrum, while the elasticity with respect to labor is relatively high.[24] Regarding IT-Outsourcing we find coefficients in the range of 0.4 - 0.6. This suggests much higher effects of IT-outsourcing on productivity than the previous literature, but results are generally hard to compare across different applications, data and implementations.[25]

Moreover, we maintained the assumption that IT outsourcing has a constant effect on (log)-productivity. This assumption is natural for a Cobb-Douglas production technology, but for a more flexible technology it is not guaranteed that the effect of IT outsourcing is linear. In such settings our specification could be extended by including interaction terms between IT outsourcing, and labor and capital inputs.

4. MONTE CARLO SIMULATION

In this section, we study the finite-sample performance of our estimator and present the results of a Monte Carlo simulation. We perform 1000 Monte Carlo replications and the sample size is equal to the size of the original data in our application, that is, $n = 2631$.

We consider the estimation of the parameter of interest β_0 under the following simulation design which mimics the empirical application. The data are generated by the binary instrument $W_2 = \mathbf{1}\{\widetilde{W}_2 > 0\}$ and the binary endogenous variable $X = \mathbf{1}\{\xi\widetilde{W}_2 + \sqrt{1-\xi^2}\widetilde{U} > 0\}$, where $\mathbf{1}\{\cdot\}$ denotes the indicator function. In addition, we have $W_1 = 3+\gamma$ and $U = \eta\widetilde{U} + \sqrt{1-\eta^2}\,\varepsilon$, where $(W_1, \widetilde{W}_2, \widetilde{U}, \varepsilon) \sim \mathcal{N}(0, I_4)$ and I_4 denotes the identity matrix of dimension four. We then draw Y^* from the model

$$Y^* = 3 + \beta_0 X + W_1 + \sigma_U U$$

[23] The assumption is restrictive and at odds with the vast majority of the literature on the estimation of production technology. Thus before using our estimation results in practice, they should be subjected to further scrutiny. In principle, the extended estimation strategy in Section 2.4 of our paper is able to accommodate more than one endogenous regressor, but at the cost of greater expositional complexity.

[24] Constant returns to scale are found in most of the previous literature in this respect (see e.g Bertschek and Kaiser, 2004; Brynjolfsson and Hitt, 2003b; Griffith et al., 2006; Ohnemus, 2009). The elasticity of productivity with respect to capital tends to range from 0.15 to 0.3 in the previous literature, and the elasticity w.r.t. labor typically ranges from 0.7 to 0.85 (Bertschek and Kaiser (2004); Brynjolfsson and Hitt (2003b) Our elasticity w.r.t. capital is 0.07, which is similar to estimates by Griffith et al. (2006). In addition, the estimates become more similar to the related when we apply our correction, but, again, it is difficult to compare these estimates, because we approximate capital with investment and cannot correct for the endogeneity in the inputs.

[25] Han et al. (2011) looked at IT outsourcing, but used data from only one economic sector. Amiti and Wei (2009) also analyzed service offshoring and looked at the first differences. Görg and Hanley (2005) found no effect of service offshoring, but focused only on the Irish electronics industry Görg et al. (2008) and Girma and Görg (2004) looked at various types of (any) outsourcing and their effects depend on the model they use to estimate the coefficients. Also own prior work (Ohnemus (2009)) found that the estimated coefficients depend considerably on the estimation procedure. Our estimates seem quite high, and a word of caution is in place.

where we let $\sigma_U = 1.5$. In our simulations, the parameter β_0 varies between 0.5 and 1. Also the parameters ξ, which measures the strength of the instrument, and η, which captures the degree of endogeneity, are varied in the experiments.

We generate realizations $Y = \Delta Y^*$ where the selection variable Δ is drawn from the Bernoulli distribution

$$\Delta \sim \textsc{Bernoulli}\Big(\Phi(Y^*/2 + \eta/2)\Big)$$

and $\eta \sim \mathcal{N}(0, 1)$ is independent from the other variables. The missingness of Y is hence selective as it depend on the partially observed outcome Y^*.

We implement the estimator $\widehat{\beta}_n$ given in (2.7). The estimator $\widehat{g}_n$ of the inverse selection probability $g(w_1, \vartheta) = \mathbf{E}(Y/G(V'\vartheta)|W_1 = w_1)$ and the estimator $\widehat{h}_n$ of the conditional expectation $h(w_1) = \mathbf{E}(X|W_1 = w_1)$ are both implemented using quadratic B-spline basis functions and five interior knots. In addition to the estimator $\widehat{\beta}_n$ we also implement an estimator which builds on the missing at random assumption. We denote $\widetilde{\beta}_n$ the estimator of β_0 which is based on listwise deletion. In our simulations, we also compute the coverage based on 500 bootstrap replications in each Monte Carlo iteration.

Table 6 shows the parameter estimates and the coverage at the 5% nominal level for the partially linear model with correction and under listwise deletion. We vary β_0 over experiments with values 0.5 and 1. The values for η are varied between 0.3 and 0.5 and for ξ between 0.5 and 0.7, respectively. From Table 6 we see that ignoring selective missingness leads to downward biased results, i.e., the MCAR estimator $\widetilde{\beta}_n$ has on average roughly a downward bias of 5% while our selection corrected estimator $\widehat{\beta}_n$ is fairly accurate. In particular, we see that the coverage based on bootstrap standard errors is very close to the true value, while there is undercoverage for the MCAR estimator $\widetilde{\beta}_n$.

5. CONCLUSIONS

Selective item non-response is a major problem in all survey-based data. We propose a novel approach to correct for potential biases in the estimation of econometric models when the dependent variable is subject to missing data. Prevalent strategies in applied empirical research to deal with missing data rely on MAR, namely, listwise deletion or multiple imputation. We show that these approaches can lead to biased estimates of the central coefficients. The bias is most likely when the missingness is related to the independent variables in systematic ways, and its sign depends on this relationship.

We develop a new estimation approach that can be used in IV estimation and is robust to selectively missing realizations of the dependent variable. The approach is based on a second set of instrumental variables that affect selection only through partially observed outcomes. We apply our proposed method to revisit the estimation of productivity returns to IT outsourcing. We argue that in such settings, that is, production function estimation based on survey data, MAR is likely violated. Our empirical application in fact supports this hypothesis. Our estimator is easily applied, and we find positive and economically meaningful productivity returns to IT outsourcing. Importantly, the positive returns are underestimated when standard methodologies are employed that assume MAR (listwise deletion and multiple imputation).

Our results highlight the consequences of the widely used MAR assumption within a broadly applied class of empirical models (production function estimation). The literature dealing with estimation of production functions has so far focused on bias due to endogenous input choice and on endogeneity through panel attrition and unit non-response (firm exit) (cf. Olley and Pakes (1996), Levinsohn and Petrin (2003), Melitz and Polanec (2015)). We highlight that, in addition, imposing MAR on missing values in the *dependent* variable is likely to yield biased estimates in this context.

Finally, our new estimator can be fruitfully used in applied empirical research with either continuous or binary instruments. We provide a semiparametric version and a version for linear IV (2SLS) of the estimator, and we show an application for the broad class of production function estimation models. However, we note that the relevance of selective missingness of the dependent variable in our application carries over to many other applications and important datasets, such as the US Census Bureau's Census of Manufacturers, the IAB establishment panel, or other firm-, individual-, and household-level surveys.

A. SUMMARY STATISTICS AND ESTIMATION TABLES

Table 1. Industry Distribution

	Obs.	Percent
consumer goods	251	9.54
chemical industry	138	5.25
other raw materials	239	9.08
metal and machine construction	309	11.74
electrical engineering	177	6.73
precision instruments	230	8.74
automobile	167	6.35
wholesale trade	135	5.13
retail trade	199	7.56
transportation & postal services	202	7.68
banks & insurances	154	5.85
technical services	230	8.74
other business-related services	200	7.60
Total	2631	99.99

Notes: This table shows the number of firms in the estimation sample by industry.
Source: ZEW ICT-Survey 2004.

Table 2. Summary Statistics: Raw Data

	ln(prod)	sales (in million Euros)	cost share	sales & cost share miss	ln(Capital)	ln(Labor)	outsource	y2k-Consult	pcwork	ost
nobs	3801	3801	3801	3801	3801	3801	3801	3801	3801	3801
NAs	1389	856	1024	491	1058	0	34	172	13	0
Mean	4.399	108.975	0.360	1.000	5.124	4.075	0.361	0.521	44.990	0.246
StD.	0.720	857.663	0.209	0.000	2.583	1.750	0.480	0.500	32.115	0.431
p50	4.306	8.000	0.350	1.000	5.017	3.951	0.000	1.000	40.000	0.000
Min	2.372	0.033	0.010	1.000	0.000	0.000	0.000	0.000	0.000	0.000
Max	8.083	29000.000	0.990	1.000	14.809	11.002	1.000	1.000	100.000	1.000

NOTES: Summary statistics of the main variables in the raw dataset. The variables are shown in the columns, the descriptive statistics in the rows. The statistic S&C NAs counts the number of observations for which total sales and the share of sales attributed to intermediate inputs and external costs (*cost share*) are both missing. Source: ZEW ICT-Survey 2004.

Table 3. Summary Statistics: Estimation Sample

	ln(prod)	sales (in million Euros)	cost share	sales & cost share miss	ln(Capital)	ln(Labor)	outsource	y2k-Consult	pcwork	ost
nobs	2631	2631	2631	2631	2631	2631	2631	2631	2631	2631
NAs	535	276	349	90	0	0	0	0	0	0
Mean	4.388	87.608	0.363	1.000	5.129	3.889	0.394	0.521	42.966	0.261
StD.	0.710	837.060	0.209	0.000	2.570	1.707	0.489	0.500	31.726	0.439
p50	4.287	6.200	0.350	1.000	5.081	3.761	0.000	1.000	33.000	0.000
Min	2.372	0.033	0.010	1.000	0.000	0.000	0.000	0.000	0.000	0.000
Max	8.083	29000.000	0.950	1.000	14.809	11.002	1.000	1.000	100.000	1.000

NOTES: Summary statistics of the main variables in the estimation sample. The variables are shown in the columns, the descriptive statistics in the rows. The statistic S&C NAs counts the number of observations for which total sales and the share of sales attributed to intermediate inputs and external costs (*cost share*) are both missing.
Source: ZEW ICT-Survey 2004.

Table 4. First Stage Results (two stage least squares)

	Dependent variable:	
	outsource	
	(1)	(2)
y2k-Consult		0.210*** (0.019)
ln(Labour)	−0.046*** (0.009)	−0.059*** (0.008)
ln(Capital)	−0.003 (0.006)	−0.002 (0.005)
pcwork	−0.002*** (0.0003)	−0.002*** (0.0003)
ost	0.044** (0.021)	0.046** (0.021)
consumer goods	−0.049 (0.046)	−0.063 (0.045)
chemical industry	−0.036 (0.054)	−0.060 (0.053)
other raw materials	−0.073 (0.047)	−0.093** (0.046)
metal and machine construction	−0.129*** (0.044)	−0.128*** (0.043)
electrical engineering	−0.237*** (0.050)	−0.220*** (0.049)
precision instruments	−0.122*** (0.047)	−0.125*** (0.045)
automobile	−0.060 (0.052)	−0.072 (0.051)
wholesale trade	0.010 (0.053)	−0.013 (0.052)
retail trade	−0.076 (0.049)	−0.099** (0.048)
transportation and postal services	−0.067 (0.049)	−0.069 (0.048)
banks and insurances	0.029 (0.051)	0.006 (0.050)
electronic processing and telecommunication	−0.192*** (0.046)	−0.176*** (0.045)
other business-related services	0.731*** (0.045)	0.675*** (0.044)
Observations	2,631	2,631
Adjusted R^2	0.055	0.098
F Statistic	10.545*** (df = 16; 2614)	17.833*** (df = 17; 2613)

Notes: This table shows the first stage results of the two stage least squares regression using the full estimation sample. The coefficient for the excluded instrument, *Y2K*, is shown in the first row. Column 1 shows estimation results without the excluded instrument and column 2 provides regression results including the instrument. Significant at 1% ***, significant at 5% **, significant at 10% *.

Table 5. Main Results

	Linear Estimation			Semiparametric Estimation		
	List. Del. (MCAR)	Imput. (MAR)	Correction	List. Del. (MCAR)	Imput. (MAR)	Correction
IT out.	0.407	0.443	0.591	0.431	0.488	0.577
SE	0.133	0.150	0.313	0.136	0.153	0.316
controls	YES	YES	YES	YES	YES	YES
N	2096	2631	2631	2096	2631	2631

NOTES: This table shows our main estimation results for the effect of IT outsourcing on productivity. The dependent variable is log(sales-costs of intermediates/employees). Standard errors are in parenthesis. The coefficient of interest is the effect of IT outsourcing on productivity. We use the "Y2K bug consulting" as the IV for IT outsourcing. Columns 1-3 show the results of the fully parametric model, using two-stage least squares and instrumenting IT outsourcing using the Y2K bug. The specifications in Columns 4-6 nonparametrically control for investment and firm size (number of employees), and parametrically for the share of employees working predominantly with personal computers, sector and location in Eastern Germany. Columns 4-6 use non-parametric series estimation. Columns 1 and 4 show the biased benchmark using listwise deletion (MCAR). Columns 2 and 5 correct for missingness using imputation methods that assume missing at random (MAR). Columns 3 and 6 apply the correction for selective non-response that we present in this paper. All standard errors are obtained using bootstrap with 500 repetitions. Significant at 1% ***, significant at 5% **, significant at 10% *.

A.1. Simulation Results

Table 6. Simulation Results

Parameters			Correction		Listw. Deletion (MCAR)	
β_0	ρ	ξ	$\widehat{\beta}_n$	coverage for β_0	$\widetilde{\beta}_n$	coverage for β_0
0.5	0.3	0.5	0.512	0.944	0.450	0.941
	0.5	0.5	0.508	0.951	0.464	0.946
	0.3	0.7	0.512	0.951	0.458	0.936
	0.5	0.7	0.511	0.948	0.464	0.933
1.0	0.3	0.5	1.003	0.949	0.912	0.914
	0.5	0.5	1.012	0.938	0.928	0.934
	0.3	0.7	1.002	0.950	0.909	0.876
	0.5	0.7	1.017	0.953	0.929	0.900

NOTES: Coverage at the 5% nominal level.

B. APPENDIX WITH PROOFS OF RESULTS

Proof of Theorem 2.1: Recall that by equation (2.4) we have

$$\beta_0 = \frac{\mathbf{E}[(Y^* - \mathbf{E}(Y^*|W_1))W_2]}{\mathbf{E}[(X - \mathbf{E}(X|W_1))W_2]}.$$

It is sufficient to consider $\mathbf{E}[(Y^* - \mathbf{E}(Y^*|W_1))W_2]$. First, we observe

$$\begin{aligned}
\mathbf{E}(Y^*|X,W) &= \mathbf{E}\left[Y^* \frac{\mathbb{P}(\Delta = 1|V^*)}{G(\vartheta_0' V^*)}\Big|X, W\right] && \text{(by Assumption 2.3 } (i))\\
&= \mathbf{E}\left[Y^* \frac{\mathbb{P}(\Delta = 1|V^*, X)}{G(\vartheta_0' V^*)}\Big|X, W\right] && \text{(by Assumption 2.2)}\\
&= \mathbf{E}\left[\mathbf{E}\left(\frac{\Delta Y^*}{G(\vartheta_0' V^*)}\Big|V^*, X\right)\Big|X, W\right]\\
&= \mathbf{E}\left[\frac{Y}{G(V'\vartheta_0)}\Big|X, W\right] && \text{(by law of total expectation)},
\end{aligned}$$

where for the last equation we used that $G(\vartheta_0' V) = G(\vartheta_0' V^*)$ whenever Δ is different from zero. In particular, by conditioning both sides of the previous conditional mean equation by W we obtain

$$\mathbf{E}(Y^*|W) = \mathbf{E}\left[\frac{Y}{G(V'\vartheta_0)}\Big|W\right].$$

This shows that after the parameter ϑ_0 is identified through the instrumental variable restriction (2.6), we can identify the conditional mean $\mathbf{E}(Y^*|W)$ through inverse probability weighting. In particular, we obtain

$$\mathbf{E}[(Y^* - \mathbf{E}(Y^*|W_1))W_2] = \mathbf{E}\left[\Big(Y/G(V'\vartheta_0) - \mathbf{E}(Y/G(V'\vartheta_0)|W_1)\Big)W_2\right]$$

which completes the proof. □

We continue this Appendix by giving the conditions under which the asymptotic distribution result summarized in Theorem 2.2 is valid.

Assumption B.1. *(i) We observe a sample* $((\Delta_1, Y_1, X_1, W_{11}, W_{21}), \dots, (\Delta_n, Y_n, X_n, W_{1n}, W_{2n}))$ *of independent and identical distributed (i.i.d.) copies of* (Δ, Y, X, W_1, W_2) *where* $Y = \Delta Y^*$. *(ii) It holds* $\sup_w \|p^L(w)\|^2 = O(L)$ *with* $L \equiv L(n)$ *and* $L^2/n = o(1)$. *(iii) The smallest eigenvalue of* $\mathbf{E}[p^L(W_1)p^L(W_1)']$ *is bounded away from zero uniformly in n. (iv) It holds* $n\,\mathbf{E}[|\gamma' p^L(W_1) - g(W_1, \vartheta_0)|^2] = o(1)$ *where* $\gamma = \mathbf{E}[g(W_1, \vartheta_0)p^L(W_1)]$. *(v) The parameter space* Θ *is compact; the function G is differentiable and* $\|G_\vartheta(v'\theta)\|$ *is bounded for every v and* $\theta \in \Theta$. *(vi) The moments* $\mathbf{E}\,|W_2\, Y/G(V'\vartheta_0)|^4$, $\mathbf{E}\,|YVG_\vartheta(V'\theta)/G^2(V'\theta)|^2$ *uniformly in* $\theta \in \Theta$, *and* $\mathbf{E}\,\|Z\|^4$ *are bounded from above.*

Assumption B.1 (*ii*) – (*iii*) restricts the magnitude of the approximating functions $\{p_j\}_{j\geq 1}$ and imposes nonsingularity of their second moment matrix. It is a standard assumption for series estimators (cf., *e.g.*, Assumption 2 in Newey (1997)). Assumption B.1 (*ii*) holds for instance for polynomial splines, Fourier series and wavelet bases. Assumption B.1 (*iv*) imposes an undersmoothing condition on the sieve approximation errors which characterize the bias of the estimated function $g(\cdot, \vartheta_0)$. This ensures that these sieve approximation biases in our estimation procedures become asymptotically negligible. In addition to this, we require smoothness of the function G.

Proof of Theorem 2.2: Due to consistency of the estimator $\widehat{h}_n$ it is sufficient to consider

$$\begin{aligned}
&n^{-1/2}\sum_{i=1}^n \Big\{W_{2i}\Big(Y_i/G(V_i'\widehat{\vartheta}_n) - \widehat{g}_n(W_{1i}, \widehat{\vartheta}_n)\Big) - \mathbf{E}\Big[W_2\Big(Y/G(V'\vartheta_0) - g(W_1, \vartheta_0)\Big)\Big]\Big\}\\
&= \underbrace{n^{-1/2}\sum_{i=1}^n \Big\{W_{2i}\Big(Y_i/G(V_i'\vartheta_0) - \gamma' p^L(W_{1i})\Big) - \mathbf{E}\Big[W_2\Big(Y/G(V'\vartheta_0) - \gamma' p^L(W_1)\Big)\Big]\Big\}}_{I}\\
&\quad \underbrace{+\, n^{-1/2}\sum_{i=1}^n Y_i\Big(W_{2i} - p^L(W_{1i})'(\mathbf{W}_n'\mathbf{W}_n)^{-1}\sum_{i'=1}^n W_{2i'}\, p^L(W_{1i'})\Big)\Big(1/G(V_i'\widehat{\vartheta}_n) - 1/G(V_i'\vartheta_0)\Big)}_{II}\\
&\quad \underbrace{+\, n^{-1/2}\sum_{i=1}^n W_{2i}\Big(\gamma' p^L(W_{1i}) - g(W_{1i}, \vartheta_0)\Big)}_{III},
\end{aligned}$$

where $\gamma = \mathbf{E}[g(W_1, \vartheta_0)p^L(W_1)]$. We further make use of the notation $\widehat{\mathbf{E}}_n[W_2|W_{1i}] =$

$p^L(W_{1i})'(\mathbf{W}_n'\mathbf{W}_n)^{-1}\sum_{i'=1}^n W_{2i'}\,p^L(W_{1i'})$. For some $\overline{\vartheta}_n$ between ϑ_0 and $\widehat{\vartheta}_n$ we have

$$\begin{aligned}
II &= \sqrt{n}(\widehat{\vartheta}_n-\vartheta_0)'n^{-1}\sum_{i=1}^n\Big(W_{2i}-\widehat{\mathbf{E}}_n[W_2|W_{1i}]\Big)Y_iV_i\,G_\vartheta(V_i'\overline{\vartheta}_n)/G^2(V_i'\overline{\vartheta}_n)\\
&= \sqrt{n}(\widehat{\vartheta}_n-\vartheta_0)'n^{-1}\sum_{i=1}^n\Big(W_{2i}-\mathbf{E}[W_2|W_{1i}]\Big)Y_iV_i\,G_\vartheta(V_i'\overline{\vartheta}_n)/G^2(V_i'\overline{\vartheta}_n)\\
&\quad+\sqrt{n}(\widehat{\vartheta}_n-\vartheta_0)'n^{-1}\sum_{i=1}^n\Big(\mathbf{E}[W_2|W_{1i}]-\widehat{\mathbf{E}}_n[W_2|W_{1i}]\Big)Y_iV_i\,G_\vartheta(V_i'\overline{\vartheta}_n)/G^2(V_i'\overline{\vartheta}_n)\\
&= \sqrt{n}(\widehat{\vartheta}_n-\vartheta_0)'\,\mathbf{E}\Big[(W_2-\mathbf{E}[W_2|W_1])YVG_\vartheta(V'\vartheta_0)/G^2(V'\vartheta_0)\Big]+o_p(1),
\end{aligned}$$

due to the the uniform law of large numbers and the $o_p(1)$ term is due to $\sqrt{n}(\widehat{\vartheta}_n-\vartheta_0)=O_p(1)$, $n^{-1}\sum_i|\,\mathbf{E}[W_2|W_{1i}]-\widehat{\mathbf{E}}_n[W_2|W_{1i}]\|^2=o_p(1)$, and the moment restrictions imposed in Assumption B.1 (*vi*). The GMM estimator $\widehat{\vartheta}_n$ of the parameter ϑ_0 is given by $\widehat{\vartheta}_n=\arg\min_{\theta\in\Theta}\|\sum_{i=1}^n Z_i\,(\Delta_i/G(V_i'\vartheta_0)-1)\|^2$ and hence standard calculation shows

$$\sqrt{n}(\widehat{\vartheta}_n-\vartheta_0)=(A'A)^{-1}A'\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i\Big(\Delta_i/G(V_i'\vartheta_0)-1\Big)+o_p(1)$$

where $A=\mathbf{E}\Big[ZV'G_\vartheta(V'\vartheta_0)/G^2(V'\vartheta_0)\Big]$. This computation yields

$$\begin{aligned}
\frac{\sqrt{n}}{\sigma}\Big(I+II\Big)&=\sum_{i=1}^n\frac{1}{\sqrt{n}\sigma}\Big(W_{2i}\Big(Y_i/G(V_i'\vartheta_0)-g(W_{1i},\vartheta_0)\Big)-\mathbf{E}\Big[W_2\Big(Y/G(V'\vartheta_0)-g(W_1,\vartheta_0)\Big)\Big]\\
&\quad-\Big(\Delta_i/G(V_i'\vartheta_0)-1\Big)Z_i'A(A'A)^{-1}\,\mathbf{E}\Big[(W_2-\mathbf{E}[W_2|W_1])YVG_\vartheta(V'\vartheta_0)/G^2(V'\vartheta_0)\Big]\Big)+o_p(1)\\
&\equiv\sum_{i=1}^n s_{in}+o_p(1).
\end{aligned}$$

Moreover, s_{in}, $1\le i\le n$, satisfy the Lindeberg conditions, which can be seen as follows. It holds $\mathbf{E}[s_{in}]=0$ and $n\,\mathbf{E}[s_{in}^2]=1$. For all $\varepsilon>0$ we observe

$$\begin{aligned}
\sum_i\mathbf{E}[s_{in}^2\mathbf{1}_{\{|s_{in}|>\varepsilon\}}]&=n\varepsilon^2\,\mathbf{E}[|s_{in}/\varepsilon|^2\mathbf{1}_{\{|s_{in}/\varepsilon|>1\}}]\\
&\le n\varepsilon^2\,\mathbf{E}\,|s_{in}/\varepsilon|^4\\
&\le Cn^{-1}\varepsilon^{-2}\Big(\mathbf{E}\Big|W_2\Big(Y/G(V'\vartheta_0)-g(W_1,\vartheta_0)\Big)\Big|^4\\
&\quad-\mathbf{E}\,\|Z\|^4\,\Big\|\,\mathbf{E}\Big[(W_2-\mathbf{E}[W_2|W_1])A(A'A)^{-1}YVG_\vartheta(V'\vartheta_0)/G^2(V'\vartheta_0)\Big]\Big\|^4\Big)\\
&=o(1).
\end{aligned}$$

The central limit theorem of Lindeberg-Feller thus implies $\sum_{i=1}^n s_{in}\xrightarrow{d}\mathcal{N}(0,1)$. Finally, since W_2 is a binary variable we obtain

$$\begin{aligned}\mathbf{E}\,|III|^2 &= n\,\mathbf{E}\left|W_2\left(\gamma' p^L(W_1) - g(W_1,\vartheta_0)\right)\right|^2 \\ &= O\Big(n\,\mathbf{E}\,|\gamma' p^L(W_1) - g(W_1,\vartheta_0)|^2\Big) \\ &= o(1),\end{aligned}$$

due to Assumption B.1 (*iv*), which implies $nIII = o_p(1)$ and hence completes the proof. □

C. APPENDIX WITH ROBUSTNESS CHECKS

A challenge to our identification strategy is a potential violation of Exclusion Restriction 2 in our application, which assumes that the need to hire y2k-consulting hits firms randomly and is not correlated with general managerial characteristics, in particular the firm's general tendency to use outsourcing, e.g. in accounting or for the general production.

While exclusion restrictions cannot be tested, we can address the concern, that y2k-consulting might be driven by the general experience and attitude towards outsourcing. Specifically, we run a robustness check on a subsample where we control for general outsourcing in addition to y2k-consulting and thus rule out this channel. To be precise, we observe an earlier wave of the survey data, that was collected in 2002 and for which the sample of interviewed firms partly overlaps with the sample of our dataset. Thus, we obtain a reduced panel of 1246 firms which are observed in both surveys, and we know whether these firms outsourced a part of their general production in the years 1999 to 2001 ("general outsourcing"). Therefore, we can use the sample of 1246 firms that were interviewed in both waves to replicate our main findings additionally controlling for general outsourcing. The result of this effort is shown in Tables C1, C2 and C3 below.

In Table C2, we run the same specification as before, but add a binary control variable for general outsourcing. Doing so confirms the main result, and highlights the problem we address with this paper even more clearly than our main specification. The coefficients for IT outsourcing remain positive and are significant at similar levels in the corrected specification, but the normal IV shows *insignificant* coefficients, when we use the reduced sample and control for general outsourcing. This difference in the uncorrected results, highlights how non-random missingness can mislead the researcher to maintaining the null hypothesis, because the coefficients are biased toward zero.

It is important to note that Table C2 not only added an additional control variable, but also reduced the sample size. We thus replicated our main specification with the same sample as in Table C2 to see which of the two factors drives the change in significance levels of the uncorrected specification. The result is shown in Table C1. The comparison shows that controlling for general outsourcing gives almost identical coefficient estimates as Table C2 and minimally tighter confidence intervals in the corrected specification. The same pattern emerges in the uncorrected specification, and also here the coefficients barely change. Finally, Table C3 shows the conditional correlation of y2K-consulting and general outsourcing, when controlling for the other variables in our specification. The correlation is very close to zero, which we take as additional evidence that y2K-consulting affected firms in a way that was orthogonal to their general inclination to use outsourcing. Moreover, after holding the general experience with outsourcing constant, our specification to measure the effect of IT-outsourcing, using y2k consulting as an instrument, takes all other tendencies towards outsourcing into account.

Replication of main findings only for panel observations

Table C1. Replication of Main Results for Panel Observations

	Listwise Deletion (MCAR)	Correction	Listwise Deletion (MCAR)	Correction
coefficient	0.142	0.942	0.125	0.940
SE	0.254	0.551	0.256	0.564
N	1004	1246	1004	1246

Notes: This table shows the regression results from the main specification, when using the same reduced sample as in Table C2, i.e. using only panel observations for which we are able to observe general outsourcing between 1999 and 2001; Source: ZEW ICT-Survey 2004 and 2002.

Robustness Check: main specification but using lagged outsourcing:

Table C2. Replication of Main Results Including General Outsourcing as Control

	Listwise Deletion (MCAR)	Correction	Listwise Deletion (MCAR)	Correction
coefficient	0.141	0.945	0.123	0.941
SE	0.234	0.551	0.240	0.564
N	1004	1246	1004	1246

Notes: This table shows the regression results from the main specification, when using panel observations and controlling for general outsourcing; Source: ZEW ICT-Survey 2004 and 2002.

Table C3. Regression: Y2K-Consulting on Lagged Outsourcing and Controls

	Dependent variable:
	y2k-Consult
$outsourcing_{t-1}$	−0.004 (0.032)
ln(Labour)	0.073*** (0.013)
ln(Capital)	−0.005 (0.008)
pcwork	0.001** (0.0005)
ost	0.015 (0.033)
consumer goods	0.123* (0.072)
chemical industry	0.063 (0.078)
other raw materials	0.137* (0.073)
metal and machine construction	−0.011 (0.068)
electrical engineering	−0.058 (0.074)
precision instruments	0.014 (0.071)
automobile	0.065 (0.082)
wholesale trade	0.162** (0.081)
retail trade	0.185** (0.075)
transportation and postal services	0.183** (0.075)
banks and insurances	0.070 (0.075)
electronic processing and telecommunication	−0.098 (0.074)
other business-related services	0.171** (0.069)
Observations	1,246
Adjusted R^2	0.067
F Statistic	6.260*** (df = 17; 1228)

REFERENCES

Abraham, K. G. and S. K. Taylor (1996). Firms' use of outside contractors: Theory and evidence. *Journal of Labor Economics 14*(3), 394–424.

Amiti, M. and J. Konings (2007). Trade liberalization, intermediate inputs, and productivity: Evidence from Indonesia. *The American Economic Review 97*(5), 1611–1638.

Amiti, M. and S.-J. Wei (2009). Service offshoring and productivity: Evidence from the United States. *The World Economy 32*(2), 203–220.

Bertschek, I. and U. Kaiser (2004). Productivity effects of organizational change: microeconometric evidence. *Management Science 50*(3), 394–404.

Bertschek, I., J. Ohnemus, and S. Viete (2017). The ZEW ICT survey 2002 to 2015: Measuring the digital transformation in german firms. *Journal of Economics and Statistics*.

Black, S. E. and L. M. Lynch (2001). How to compete: the impact of workplace practices and information technology on productivity. *Review of Economics and Statistics 83*(3), 434–445.

Bloom, N., R. Sadun, and J. Van Reenen (2012). Americans do IT better: Us multinationals and the productivity miracle. *The American Economic Review 102*(1), 167–201.

Borgoni, R., D. Marasini, and P. Quatto (2012). Handling nonresponse in business surveys. In *Survey Research Methods*, Volume 6, pp. 145–154.

Bresnahan, T. F., E. Brynjolfsson, and L. M. Hitt (2002). Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *The Quarterly Journal of Economics 117*(1), 339–376.

Breunig, C. (2017). Testing missing at random using instrumental variables. *Journal of Business & Economic Statistics*, 1–12.

Breunig, C., E. Mammen, and A. Simoni (2018). Nonparametric estimation in case of endogenous selection. *Journal of Econometrics 202*(2), 268–285.

Brynjolfsson, E. and L. M. Hitt (2003a). Computing productivity: Firm-level evidence. *Review of economics and statistics 85*(4), 793–808.

Brynjolfsson, E. and L. M. Hitt (2003b). Computing productivity: Firm-level evidence. *The Review of Economics and Statistics 85*(4), 793–808.

Coase, R. H. (1937). The nature of the firm. *Economica 4*(16), 386–405.

Das, M., W. K. Newey, and F. Vella (2003). Nonparametric estimation of sample selection models. *The Review of Economic Studies 70*(1), 33–58.

D'Haultfoeuille, X. (2010). A new instrumental method for dealing with endogenous selection. *Journal of Econometrics 154*(1), 1–15.

Drechsler, J. (2010). Multiple imputation of missing values in the wave 2007 of the IAB establishment panel. Technical report, IAB discussion paper.

Earp, M., M. Mitchell, J. McCarthy, and F. Kreuter (2014). Modeling nonresponse in establishment surveys: Using an ensemble tree model to create nonresponse propensity scores and detect potential bias in an agricultural survey. *Journal of Official Statistics 30*(4), 701–719.

Feenstra, R. C. (1998). Integration of trade and disintegration of production in the global economy. *The Journal of Economic Perspectives*, 31–50.

Foster, L., J. Haltiwanger, and C. Syverson (2008). Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *The American Economic Review 98*(1), 394–425.

Frick, J. R. and M. M. Grabka (2010). Item non-response and imputation of annual labor income in panel surveys from a cross-national perspective. *Survey Methods in multicultural, multinational, and multiregional Contexts. Wiley, New York.*

Girma, S. and H. Görg (2004). Outsourcing, foreign ownership, and productivity: evidence from uk establishment-level data. *Review of International Economics 12*(5), 817–832.

Görg, H. and A. Hanley (2005). International outsourcing and productivity: evidence from the irish electronics industry. *The North American Journal of Economics and Finance 16*(2), 255–269.

Görg, H., A. Hanley, and E. Strobl (2008). Productivity effects of international outsourcing: evidence from plant-level data. *Canadian Journal of Economics/Revue canadienne d'Économique 41*(2), 670–688.

Griffith, R., E. Huergo, J. Mairesse, and B. Peters (2006). Innovation and productivity across four european countries. *Oxford review of economic policy 22*(4), 483–498.

Grossman, G. M. and E. Helpman (2005). Outsourcing in a global economy. *Review of Economic Studies 1*(72), 135–159.

Grossman, S. J. and O. D. Hart (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *The Journal of Political Economy 4*(94), 691–719.

Halpern, L., M. Koren, and A. Szeidl (2015). Imported inputs and productivity. *The American Economic Review 105*(12), 3660–3703.

Han, K., R. J. Kauffman, and B. R. Nault (2011). Returns to information technology outsourcing. *Information Systems Research 22*(4), 8241–7840.

Heckman, J. (1974). Shadow prices, market wages, and labor supply. *Econometrica: Journal of the Econometric Society*, 679–694.

Kennickell, A. B. (1998). Multiple imputation in the survey of consumer finances. In *Proceedings of the Section on Survey Research Methods*.

Kleibergen, F. and R. Paap (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of econometrics 133*(1), 97–126.

Kline, P. and A. Santos (2013). Sensitivity to missing data assumptions: Theory and an evaluation of the US wage structure. *Quantitative Economics 4*(2), 231–267.

Lacity, M. C., S. A. Khan, and L. P. Willcocks (2009). A review of the IT outsourcing literature: Insights for practice. *The Journal of Strategic Information Systems 18*(3), 130–146.

Lee, J., J. S. McCullough, and R. J. Town (2013). The impact of health information technology on hospital productivity. *The RAND Journal of Economics 44*(3), 545–568.

Levinsohn, J. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies 70*(2), 317–341.

Loh, L. and N. Venkatraman (1992). Diffusion of information technology outsourcing: influence sources and the kodak effect. *Information Systems Research 3*(4), 334–358.

Melitz, M. J. and S. Polanec (2015). Dynamic olley-pakes productivity decomposition with entry and exit. *The RAND Journal of Economics 46*(2), 362–375.

Menon, N. M., B. Lee, and L. Eldenburg (2000). Productivity of information systems in the healthcare industry. *Information Systems Research 11*(1), 83–92.

Miao, W., E. T. Tchetgen, and Z. Geng (2015). Identification and doubly robust estimation of data missing not at random with a shadow variable. *arXiv preprint arXiv:1509.02556*.

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics 79*(1), 147 – 168.

Ohnemus, J. (2007). Does IT outsourcing increase firm success? an empirical assessment using firm-level data. *ZEW Discussion Paper 07–087*.

Ohnemus, J. (2009). Productivity effects of business process outsourcing-a firm-level investigation based on panel data. *ZEW-Centre for European Economic Research Discussion Paper* (09-088).

Olley, G. S. and A. Pakes (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica 64*(6), 1263–1297.

Paiva, T. and J. Reiter (2014). Using imputation techniques to evaluate stopping rules in adaptive survey designs. *US Census Bureau Center for Economic Studies Paper No. CES-WP- 14-40*.

Petrin, A. and J. Levinsohn (2012). Measuring aggregate productivity growth using plant-level data. *The RAND Journal of Economics 43*(4), 705–725.

Ramalho, E. A. and R. J. Smith (2013). Discrete choice non-response. *The Review of Economic Studies 80*(1), 343–364.

Raymond, W., J. Mairesse, P. Mohnen, and F. Palm (2015). Dynamic models of r & d, innovation and productivity: Panel data evidence for dutch and french manufacturing. *European Economic Review 78*, 285–306.

Reardon, J., R. Hasty, and B. Coe (1997). The effect of information technology on productivity in retailing. *Journal of retailing 72*(4), 445–461.

Rubin, D. B. (1976). Inference and missing data. *Biometrika 63*, 581–592.

Rubin, D. B. (1978). Multiple imputations in sample surveys-a phenomenological bayesian approach to nonresponse. In *Proceedings of the Survey Research Methods Section of the American Statistical Association*, Volume 1, pp. 20–34. American Statistical Association.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys.* New York: Wiley & Sons.

Schreyer, P. and D. Pilat (2001). Measuring productivity. *OECD Economic studies 33*(2), 127–170.

Tang, G., R. J. Little, and T. E. Raghunathan (2003). Analysis of multivariate missing data with nonignorable nonresponse. *Biometrika 90*(4), 747–764.

Thompson, K. J. and K. T. Washington (2013). Challenges in the treatment of unit nonresponse for selected business surveys: a case study. *Survey Methods: Insights from the Field (SMIF)*.

Tomaskovic-Devey, D., J. Leiter, and S. Thompson (1994). Organizational survey nonresponse. *Administrative Science Quarterly*, 439–457.

U.S. Department of Commerce (1999). *The Economics of Y2K and the Impact on the United States*. Washington, D.C.

White, T. K., J. P. Reiter, and A. Petrin (2012). Plant-level productivity and imputation of missing data in us census manufacturing data. NBER Working Paper 17816, National Bureau of Economic Research.

Williamson, O. E. (1979). Transaction-cost economics: The governance of contractual relations. *Journal of Law and Economics 2*(22), 233–261.

Williamson, O. E. (1981). The economics of organization: The transaction cost approach. *American Journal of Sociology 3*(87), 548–577.

Williamson, O. E. (1989). *Transaction cost economics*, Volume 1, Chapter Handbook of Industrial Organization, pp. 135–182. North-Holland.

Wooldridge, J. M. (2009). On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters 104*(3), 112–114.

ZEW (2010). *Interaktiv, mobil, international – Unternehmen im Zeitalter von Web 2.0, IKT-Report*. Mannheim: Zentrum für Europäische Wirtschaftsforschung (ZEW).

Zhao, J. and J. Shao (2015). Semiparametric pseudo-likelihoods in generalized linear models with nonignorable missing data. *Journal of the American Statistical Association 110*(512), 1577–1590.