Three Essays on Econometrics and Economics of Education

by

Zizhong Yan

A Thesis Submitted to the
University of Warwick
for the Degree of

Doctor of Philosophy in Economics

Department of Economics
University of Warwick
March 2017

"The Serious Snow 𝄞"

Abstract

This dissertation is a collection of three independent essays on econometrics and economics of education. The first chapter investigates how the Research Excellence Framework (REF) perceives the quality of economics journals. Exploiting aggregate information available in the published REF dataset, we propose a novel algorithm within an ordered probit framework that provides an effective recovery of underlying disaggregated outcomes (i.e. individual submission). The estimated results can be viewed as a directory for predicting the perception of journal quality for the REF 2014 exercise. In the second chapter, I suggest a new matching methodology—the Dirichlet process (DP) matching—that has several important advantages compared to conventional matching methods, including the balancing property, a more efficient ATT estimator and a credible confidence band. I describe the DP matching as a story of the "Chinese restaurant with invited guests". In the third chapter, we exploit a quasi-natural experiment, namely the *PelCa* program in Ecuador, and study its consequences on mothers' empowerment and children's early development. We find optimistic evidence that the program helped mothers rear their children in a more learning conductive environment, resulting in positive effects on children outcomes and greater empowerment in mothers at home and in the community.

Contents

List of Figures

List of Tables

Declaration of Authorship

I herewith declare that the thesis entitled *Three Essays on Econometrics and Economics of Education* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. Chapter 1 is a joint work with Professor Michael Pitt, who is my PhD supervisor. Chapter 3 is co-authored with Professor Victor Lavy and Dr Giulia Lotti. For Chapter 1, I developed the methodologies and research ideas together with Professor Michael Pitt. Under his supervision, I programmed the econometric algorithm, prepared and analysed both simulated and actual data, and prepared the results and drafts. For Chapter 3, I joined the project during the economic and econometric analysis stage in March 2014. My involvement with this paper includes processing data, conducting econometric analysis using statistical packages, econometric modelling, helping to prepare results and drafts in relation to empirical and methodology parts of the study. I have acknowledged all main sources of help and clearly given the sources I quoted or consulted from the work of others. I confirm that this thesis has not been presented in an identical or similar form to any other examination board.

Acknowledgements

First and foremost I want to express my huge gratitude to my remarkable supervisors Michael Pitt, Sascha Becker, and Mingli Chen, for it is without whom the completion of this dissertation would not have been possible. Their immense knowledge and tremendous support are invaluable for my PhD journey.

I am very grateful to Jeremy Smith and Victor Lavy for stimulating conversations and their continuing supports on my academic development.

I thank Wiji Arulampalam, Gianna Boero, Fernanda Brollo, François Caron, Clément de Chaisemartin, Valentina Corradi, Mirko Draca, Jing Gao, Corrado Giulietti, Clément Imbert, Arthur Lewbel, Robin Naylor, Motty Perry, Roland Rathlot, Mark Steel, and Massimiliano Tani for sharing their numerous insights with me.

My friends and colleagues, especially Giulia Lotti, Hatsuyo Shimazaki, William Allen, Xiaocheng Hu, Vaclav Zdarek, Piu Yan, Mohamed Ali, Wentao Fu, Terry Cheung, and Fei Ding were a source of constructive discussions and solid support during these years.

I acknowledge the Department of Economics at the University of Warwick for materialising this thesis technically and financially. The detailed acknowledgements to each paper are provided at the beginning of each chapter.

Lastly, I thank my family for their warmth, prayers, encouragement and financial assistance, and my wife, Chuhong Wang, for her unwavering belief in me, unfading love and affection.

Introduction

This thesis comprises three independent essays in the fields of econometrics and economics of educations, documented in Chapter 1, 2 and 3. The structure of the thesis is illustrated in Figure 1.

In particular, Chapter 1 investigates how the Research Excellence Framework (REF), last held in 2014 to assess the research quality in British higher education institutions over the period 2008-2013, perceive economics journals in their assessment system. Exploiting on-line published data on submitted research outputs of different REF quality standards, which is only available at the institutional level, we propose a novel algorithm within an ordered probit framework that allows us to distinguish the censored REF standards for each individual submission and to estimate how economics journals were perceived by the Economics and Econometrics sub-panel and the Business and Management Studies sub-panel. In particular, we develop an efficient Markov Chain Monte Carlo (MCMC) sampling scheme for the inference and also suggest a robust and weakly informative prior distribution to overcome the potential separation problem. This is the first paper to employ a standard regression model to directly predict the perception of journal quality for the REF 2014 exercise. The estimated results can be viewed as a directory for determining to what extent each economics journal meets the criteria set by the REF 2014. Our proposed method can be generalised to other generalised linear models where the outcomes are censored at an aggregate level.

Chapter 2 focuses on the estimation of the average treatment effect in evaluation studies. There have been concerns that 1) conventional matching methods such as the propensity score matching cannot ensure the balancing property of the matched pairs; 2) the estimator for the average treatment effect on the treated (ATT) could be more efficient when a theoretical full covariates matching is employed; however, full covariates matching is not typically feasible in practice; and 3) two-step matching estimators are practically feasible but may result in incredibly large confidence intervals of the ATT estimator. The Dirichlet process (DP) matching strategy proposed in this paper is committed to delivering a desirable ATT estimator to address these concerns. The algorithm meets the balancing property by construction. As a compromise of full covariates matching, the DP matching effectively matches the controlled to the treated if their confounders are in the same covariate space. The whole algorithm is integrated into a single efficient Markov Chain Monte Carlo (MCMC) scheme and the resulting standard deviations can be viewed as an honest confidence band for the ATT estimator. I illustrate the usefulness of the proposed method with an empirical application of the well-known LaLonde (1986) data. I demonstrate that the DP matching ATT estimator is efficient and closer to the benchmark values than popular matching approaches.

Chapter 3 empirically evaluates an NGO preschool program in Ecuador. Empowering women and enhancing children's early development are two important goals that are often pursued via independent policy initiatives in developing countries. In this paper we study a unique approach that pursues both goals at the same time: empowering mothers through tools that also advance their children's development. A program operated by AVSI, an Italian NGO, in a poor neighborhood of Quito, Ecuador, targets parents of children from birth to age 5. It provides family advisor-guided parent training sessions once every two weeks for groups of six to eight mothers and their children. We find that the program empowered women in various dimensions: treated mothers are more likely to be employed, more of them have a full-time job and they are more likely to have a formal-sector job. They also earn higher wages, and are more likely to make independent decisions regarding their own sources of money, work outside the home, and continuing their education. Moreover, there is evidence of a greater role overall in intra-household decisions, especially on issues involving children's education and discipline. Treated mothers also increase parental inputs into their children's development, who are less likely to repeat a grade or temporarily drop-

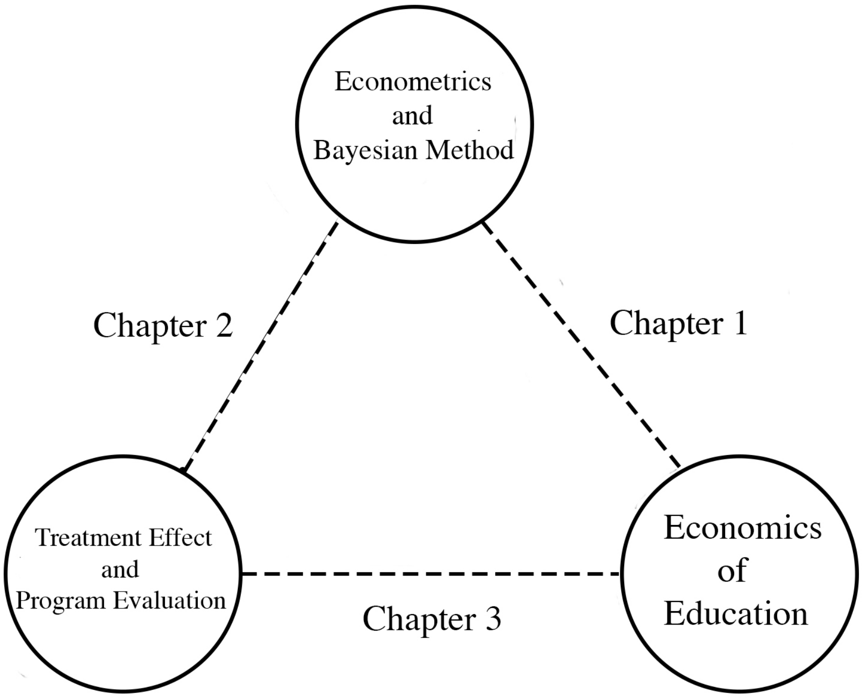


Figure 1: Thesis Structure
Chapter 1 and 2 both contribute to the econometrics and Bayesian econometrics in particular. One important feature that Chapter 1 and Chapter 3 share is their findings both hold paramount importance in the design of optimal education policy. Chapter 2 and Chapter 3 are both associated with estimating the causal effect.

out from schooling, and have higher scores on cognitive tests.

Chapter 1

How Does the REF Panel Perceive Journals? A New Approach to Estimating Ordinal Response Model with Censored Outcomes [1]

1.1 Introduction

Research quality has been an important issue for researchers and policymakers in terms of various decisions such as promotion and public investment in research. The Research Excellence Framework (REF) is aimed at providing a dynamic and internationally competitive UK research sector, and was designed to assess the research quality in British higher education institutions over the period 2008-2013. The most recent REF exercise was held in 2014. The most important task of the REF 2014 is to evaluate the quality and impact of submitted research outputs from different UK institutions on different subjects. In December 2014, the REF published the assessment results for each institution in the form of aggregate proportions of submissions in a given institution that meet four quality profiles, namely 4*, 3*, 2* and 1* (higher number indicates better quality). By design, the REF evaluates the research attainment of institutions rather than that of researchers,

[1] The authors thank Wiji Arulampalam, Sascha O. Becker, Gianna Boero, François Caron, Mingli Chen, Victor Lavy, Robin Naylor, Jeremy Smith, and Motty Perry for valuable comments. Zizhong Yan would like to thank the Department of Economics at the University of Warwick for the financial support.

and therefore the assessment outcome for each individual submission meeting a specific standard is actually censored and unobservable by the public. However, it is important for universities and researchers to know how their papers have been judged by the REF system. In some subject areas, it is possible that an automated procedure may replicate the laboratory REF exercise.

In particular, we seek to answer how economics journals or other types of submission (e.g. books, monographs and working papers) were perceived by the REF panel. In an ideal case, one can directly answer this question by running an ordinal regression of the quality outcome of each submission on journal indicators and dummies for other types of publication. Unfortunately, for the REF 2014 exercise, the only information that is available about a particular submission (e.g. Economics at the University of Leicester) is how many of their publications were rated as 4*, 3*, 2* or 1*. This aggregate data is summarised in Table 1.1. We do not know, however, in which of the four categories the Leicester Economics department's 80 papers were submitted.

In fact, this is a common problem in applied economics research—when the objective is to understand the impact of an individual-level variable, when the outcome of interest is only available at certain higher levels due to data limitations. This can be because the outcome measure is published as aggregate statistics (e.g. the REF 2014), or possibly because the outcome variable must be extracted from another data source.[2] Under such circumstances, a regression of higher level outcomes on lower level covariates can solely identify the impact on higher level responses, but reveal little information about the impact at a more local level.

The aim of this paper is to find a solution to reveal this "hidden information". We develop a novel algorithm within the ordinal response model framework to probabilistically recover the individual values of the outcomes based on aggregate information at the institutional level. Specifically, our algorithm samples the unobservable outcomes by configuring

[2]For instance, labour economists often try to understand the effect of wage incomes on people's health status based on labour force surveys. Information on wages is provided by the survey data while some of the objective health measures can only be acquired from an external dataset of medical examinations. In order to concatenate the outcomes and the covariates, one needs to stratify the health outcomes by group, such as age or occupational sub-groups. The two datasets are then merged based on these sub-groups.

their conditional posterior densities, and then exchanging their values accordingly. The exchange of probabilities is conducted for observations at the same level (institutions) so that the aggregate counts in each institution are still the same. The inference of our model is integrated into a single MCMC sampling. The conditional posterior of the outcome values is able to converge to the equilibrium distribution efficiently. For parameters of the ordered probit model, we propose an efficient Metropolis-Hasting with Laplace approximation. It is worth mentioning that in our case the performance of the MCMC sampler is more stable than the Gibbs sampling with a data augmentation procedure by Albert and Chib (1993), which is a popular choice in sampling posteriors of a probit or logit model.[3]

In the ordered response model with many dummy variables, a by-product of our algorithm pertains to the hidden collinearity. This happens when the patterns of the RHS dummy variables, and those of the LHS ordinal outcomes, are highly correlated. This is known as the *separation* problem in the statistics literature.[4] As a consequence of separation, the likelihood function is monotonically increasing, and the conventional maximum likelihood estimator is unable to reach the maxima. Heinze (2006) and Heinze and Schemper (2002) have attempted to overcome the separation in a binary logit model by adopting the Jeffreys prior, which is originally used to correct the small sample bias of maximum likelihood estimators in generalised linear models (Firth, 1993). However, the attempts made by Gelman et al. (2008) and our paper all reach the same conclusion that the Jeffreys prior cannot work stably when correcting the separation problem. As an alternative, we adopt an independent Cauchy prior with scale 10 to handle the separation issue in our ordered probit model, inspired by Gelman et al. (2008) who use the Cauchy prior with scale 4 for a binary logit model. We extend their justifications to the ordered probit context. What is important is that in most applied economics studies, the researchers are only interested in the marginal effects in a probit or logit model instead of the parameters of the regression model. Based on such preferences, we are able to justify the choice of our

[3] Dunson (2008) also suggests that, for ordinal response models, a better approach is to use the Metropolis-Hastings after marginalising out the augmented data.

[4] Zorn (2005) provides a technical review of the separation problem. In addition, a review with real examples can be found at the following website of Stata package: http://www.stata.com/support/faqs/statistics/completely-determined-in-logistic-regression/.

prior distribution.

We use the REF 2014 dataset to show that our approach can effectively restore the individual outcomes, as the predicted GPA score of each institution using our method is very similar to the actual GPA score in the real dataset. To verify the proposed algorithm and the model further, we implement four simulations using synthetic datasets that are designed to mimic the true REF data. Additionally, motivated by ideas from the machine learning literature, we also evaluate the quality of model predictions. Empirically, the paper most closely related to ours is the study by Hole (2015), who ranks economics journals by using the same REF 2014 data. His algorithm, built on correlations between submissions, generates scores for each economics journal. In comparison, the methodology in this paper does not include any ad-hoc steps, and is set strictly within a standard regression framework. Besides the point estimators, we also provide the confidence bands of the estimators.

Our empirical results reveal several interesting findings. First, as expected, famous journals such as the AER, Econometrica, and QJE are ranked at the very top, whereas the Economics Letters and other types of publication are ranked at the bottom when looking at journals with more than 20 submissions. Second, after accounting for journals with a relatively small number of submissions, we find that some *new* journals (e.g. the Quantitative Economics and the American Economic Journal series), and *qualitative* journals (e.g. the Journal of Economic History and the Explorations in Economic History) have notably high ranks. Third, we observe institution fixed effects on journal perception. For the same type of outputs published in the same journal, submissions from UCL, Cambridge, Warwick and Oxford are perceived more positively (i.e. above the average), whereas submissions from Brunel, Kent and Aberdeen appear to be perceived less positively. Last, we find that the perception of journal quality differs across different REF sub-panels.

The remainder of this paper unfolds as follows. In Section 1.2, we present an econometric model, and summarise the MCMC sampling scheme. Section 1.3 shows different simulations based on an artificial dataset in order to verify the proposed algorithm, and to screen suitable choices of the link function and the prior distribution for this paper. Section

1.4 first describes the REF 2014 dataset used in the empirical analysis, and then applies the algorithm for recovering the outcome variable, and presents the empirical results of the MCMC samplings. Section 1.5 concludes this paper.

1.2 Model and Algorithm

We now formulate our empirical investigation within an ordered probit model framework in Section 1.2.1. In Section 1.2.2, the algorithm for predicting the censored outcome of the REF quality standards is developed under an ordered probit specification. The algorithm can be viewed as a step in the Gibbs' cycle and therefore, the censored outcome can be efficiently sampled via the Gibbs sampling. Section 1.2.3-1.2.4 illustrate the prior specifications for the ordered probit model in our case. Section 1.2.5 summarises the MCMC sampling scheme.

1.2.1 Notation and Framework

Let the individual output (corresponding to each individual submission to the REF panel) be labelled as i and the institution (e.g. the Department of Economics at LSE) as j. $D_{i,k}$ $(k = 1, ..., K)$ denotes an array of dummy variables for journals or other types of submission. For simplicity of notation, at this stage we focus on an individual outcome Y_i. We will examine the full model, including the institution index j in Section 1.2.2. The outcome variable is captured by Y_i, indicating ordinal categories for the four REF quality profiles 4*, 3*, 2* and 1*[5] (see Section 1.4.1 and Table 1.1 for details). We employ an ordered probit regression model to examine how each journal, or other type of publication (D_k), is perceived according to the REF criteria (Y). Using the latent variable representation, we specify the model as follows:

$$Y_i^* = \alpha_0 + \beta' \mathbf{D}_i + \epsilon_i, \qquad \epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2) \tag{1.2.1}$$

[5]In addition to these four classes, a small number of the output that falls below 1* or does not meet the published standard was rated as "unclassified". We merge the "unclassified" category with class 1*.

where i indices the individual output, $\mathbf{D}_i = (D_{i,1}, \ldots, D_{i,K})'$ is a $K \times 1$ vector of dummy covariates for the K types of journals and other submissions, and $\beta = (\beta_1, \ldots, \beta_K)'$ is a vector of parameters for $\mathbf{D}_i$. Y^* is a latent Gaussian variable in the ordered probit model, and can be stratified into four parts using three cut-off points. The ordinal response Y_i can thus be defined accordingly:

$$Y_i = \begin{cases} 1 \text{ if } Y_i^* < c_1 \\ 2 \text{ if } c_1 \leq Y_i^* < c_2 \\ 3 \text{ if } c_2 \leq Y_i^* < c_3 \\ 4 \text{ if } Y_i^* \geq c_3 \end{cases}$$

where c_1, c_2 and c_3 are the cut-off points of the ordered probit model.

In order to identify the parameters of the ordinal response regression model and the predicted probabilities, we impose the following normalisations: 1) we assume that $Var(\epsilon_i) = 1$ and $c_s > c_{s-1}$ to preserve positive probabilities; 2) we eliminate the constant term α_0. This is because the identification of the model requires us to either drop the intercept or fix one cut-off point. As a result of the absence of the constant term, the actual cut-off points are left shifted by α_0; and finally, 3) we further force the cut-off point c_1 to be zero, which allows us to compute the probabilities of an individual output Y_i being considered to be of quality 4*, 3*, 2* or 1*. Therefore, for an individual output Y_i published at journal k (i.e. $D_{i,k} = 1$ and $D_{i,-k} = 0$), the probabilities of being ranked as 4*, 3*, 2* or 1* can be expressed as:

$$\begin{aligned} \Pr(Y_i = s) &= \Phi(c_s - \beta'\mathbf{D}_i) - \Phi(c_{s-1} - \beta'\mathbf{D}_i) \\ &= \Phi(c_s - \beta_k) - \Phi(c_{s-1} - \beta_k) \qquad \text{for} \quad s = 4^*, 3^*, 2^*, 1^* \qquad (1.2.2) \end{aligned}$$

where $c_0 = -\infty, c_1 = 0$ and $c_4 = \infty$; $\Phi(.)$ indicates the cumulative distribution function of the standard Normal distribution. In addition, it is common to apply GPA scores to epitomise the REF assessment results of each institution (see for example, the REF 2014

Table of Excellence by *Times Higher Education* (Jump, 2014)). Hence, we predict the GPA score of each institution j by:

$$\widehat{GPA_j} := \sum_{i=1}^{n_j} \sum_{s=1}^{4} \hat{\Pr}(Y_i = s) \times s$$

and similarly the predicted GPA score to represent the perception of quality for each journal k can be written as:

$$\widehat{GPA_k} := \frac{1}{4} \sum_{s=1}^{4} \hat{\Pr}(Y_i = s | D_{i,k} = 1, D_{i,-k} = 0) \times s$$

1.2.2 Exchange Probabilities to Predict Censored Outcomes

In our ordered probit model the actual ordinal response Y is, however, unobservable since the REF 2014 published data does not tell us the detailed classification of each submission. Instead, the REF provides aggregate statistics for submissions that meet the REF quality criteria, at the level of the institution (See Table 1.1). Algebraically, we are able to produce an aggregate measure for the total number of outputs T_{js} in each quality class $s, (s = 1, 2, 3, 4)$ at each institution $j, (j = 1, ..., J)$:

$$T_{js} := \sum_{i=1}^{n_j} I(Y_i = s) \qquad \text{for} \quad s = 4, 3, 2, 1 \tag{1.2.3}$$

where n_j is the total number of submissions from the institution j.

Our proposed Gibbs step for sampling Y is built upon observable aggregate measures T_{js} within the ordered probit regression framework, and provides an effective recovery of the disaggregate outcomes. Since we already know the aggregate counts of the outcome, we need make sure that the recovered disaggregate outcomes Y will not violate these aggregate counts T_{js}. Motivated by this rule, we propose the *"exchange of probabilities"* method to sample Y: given the current state of $\theta = (\beta, c)$ and Y, we swap every two distinct values of Y_i and Y_l in the same the institution (i.e. y_i and y_l, where $i \neq l$), and then assign y_i and y_l

to the configuration with the highest posterior probability. Importantly, because the values can only be swapped within the same institution, the observable aggregate measures T_{js} for all j and all s will not be changed. More specifically, we consider each output Y_i, and randomly pick another output Y_l from the same institution. In doing so, we obtain two configurations:

$$\Pr(Y_i = y_i, Y_l = y_l|\theta) \text{ and } \Pr(Y_i = y_l, Y_l = y_i|\theta) \tag{1.2.4}$$

and the joint probability of submissions Y_i and Y_l which are equal to their current values:

$$\begin{aligned} \Pr(Y_i = y_i, Y_l = y_l|\theta) &= \Pr(Y_i = y_i|\theta)\Pr(Y_l = y_l|\theta) \\ &= \Bigg(\Phi(c_{y_i} - \beta'\mathbf{D}_i) - \Phi(c_{y_i-1} - \beta'\mathbf{D}_i)\Bigg) \\ &\quad \times\Bigg(\Phi(c_{y_l} - \beta'\mathbf{D}_l) - \Phi(c_{y_l-1} - \beta'\mathbf{D}_l)\Bigg) \end{aligned} \tag{1.2.5}$$

and the joint probability that Y_i and Y_l exchange their values is:[6]

$$\begin{aligned} \Pr(Y_i = y_l, Y_l = y_i|\theta) &= \Pr(Y_i = y_l|\theta)\Pr(Y_l = y_i|\theta) \\ &= \Bigg(\Phi(c_{y_l} - \beta'\mathbf{D}_i) - \Phi(c_{y_l-1} - \beta'\mathbf{D}_i)\Bigg) \\ &\quad \times\Bigg(\Phi(c_{y_i} - \beta'\mathbf{D}_l) - \Phi(c_{y_i-1} - \beta'\mathbf{D}_l)\Bigg) \end{aligned} \tag{1.2.6}$$

Once the above two configurations have been computed, the Gibbs step of sampling

[6] The latent variable Y^* can be integrated out by:

$$\begin{aligned} f(Y_i|\theta) &= \int f(Y_i, Y_i^*|\theta)\mathrm{d}Y_i^* \\ &= \int \Pr(Y_i|Y_i^*, \theta) f_{\mathcal{N}}(Y_i^*|\theta)\mathrm{d}Y_i^* \\ &= \int I(c_{Y_i-1} < Y_i^* \leq c_{Y_i}) f_{\mathcal{N}}(Y_i^*|\theta)\mathrm{d}Y_i^* \\ &= F_{\mathcal{N}}(c_{Y_i}|\beta'\mathbf{D}_i, 1) - F_{\mathcal{N}}(c_{Y_i-1}|\beta'\mathbf{D}_i, 1) \\ &= \Phi(c_{Y_i} - \beta'\mathbf{D}_i) - \Phi(c_{Y_i-1} - \beta'\mathbf{D}_i) \end{aligned}$$

where $f_{\mathcal{N}}$ and Φ refer to the standard Normal density and its distribution function, respectively.

Y_i can be formulated.[7] First, given the current state of $\theta = (\beta, c)$, we randomly pick two indices i and l from the same institution j. Second, we compute two configurations according to Equation 1.2.5 and 1.2.6. Third, we normalise two configurations and compute the probability of swapping their values, i.e.:

$$\mathcal{P}_{il} := \frac{\Pr(Y_i = y_l, Y_l = y_i|\theta)}{\Pr(Y_i = y_i, Y_l = y_l|\theta) + \Pr(Y_i = y_l, Y_l = y_i|\theta)} \tag{1.2.7}$$

Finally, y_i and y_l swap values with a probability $\mathcal{P}_{il}$, and both stay at their current values otherwise. In each Gibbs sampling iteration, we perform the above steps to exchange the probabilities for every pair of the observations until all have been considered.

In this model, the outcome values of submissions can be exchanged based on their posterior probabilities in ways that minimise the risk of incorrect assignments. The probabilistic exchange is performed for submissions from the same department j, which guarantees that the total count T_{js} is not violated. One might want to exchange more values at one time; however, this could be computationally costly as it increases the number of configurations geometrically. In Section 1.3.1 we utilise an artificial dataset to demonstrate the validity of this algorithm.

1.2.3 Separation Problem and Prior Specification

Monotonic Likelihood Function

In a generalised linear model, such as a probit or logit regression, the separation problem often arises when the corresponding outcome observations of the covariates are perfectly or nearly perfectly determined. This problem could be more likely to occur in our ordinal regression specification where many journal dummies are simultaneously present. Indeed, a by-product of our algorithm for exchanging probabilities is the separation problem. For instance, after a few MCMC iterations with the exchange of probabilities, the journal dummy for Econometrica could have all values of 4 in the LHS variable. The separation results in a monotonic likelihood function, and, as a consequence, the normal maximum

[7] Section 1.2.5 carefully summarises the Gibbs sampling scheme for our model.

likelihood estimator will be either infinitely large or infinitely small.[8] In the Bayesian context, if one chooses a flat prior, then a very large (or very small) candidate parameter would be very easy to be accepted by the Metropolis-Hastings algorithm. Additionally, the separation could be more of a concern in our case, in which a large number of journal indicators are included.

An important piece of Bayesian literature by Firth (1993) proposes a solution to correct the well-known small sample bias from the maximum likelihood estimator for generalised linear models. In a Bayesian interpretation, this is known as the Jeffreys prior. One benefit of this method is that it can be adopted to address the *separation* problem in a logit model. However, based on our attempts, the Jeffreys prior induces the convergence problem when dealing with the separation problem. Gelman et al. (2008) has also encountered the same convergence problem. In our ordered probit set-up, the Jeffreys prior could actually be more computationally complicated compared to that of a simple binary logit model, and the MCMC updating will be inefficient in practice.

Alternatively, Gelman et al. (2008) put forward the Cauchy prior with scale 4 to handle the separation issue in a binary logit model. They then provide evidence on cross-validation to support this prior. Motivated by this idea, we extend the Cauchy prior to an ordered probit model with censored outcomes. One particular concern is then to select a prior that can produce proper predicted probabilities and marginal effects rather than focusing solely on model parameters (i.e. β and c).

1.2.4 Cauchy Prior with Scale 10

To find a suitable prior distribution for our model, we reconcile two crucial points: 1) whether the prior is strong enough to correct the monotonicity of the likelihood in the presence of perfectly determined covariates; and 2) whether the prior is diffuse enough so that it has the smallest impact on the resulting posterior parameters of both separated and non-separated covariates, and can thus deliver proper predicted probabilities and marginal

[8] Some popular statistical packages, such as R and Stata, will randomly generate a large point estimator of perfectly determined covariates as well as wide confidence intervals.

effects.

In a probit model with a standard normal error term, the coefficient on a dummy variable or a normalised continuous variable is most likely to fall within the range of [-3, 3] (i.e. within three standard deviations of the standard normal distribution). In the presence of the separation, the coefficient may become either infinitely large or small, and at the same time the predicted probabilities for any non-zero value of the separated covariate and its marginal effects will be at their extreme values. However, since the quantities of interest in most probit or logit regressions are the predicted probabilities and marginal effects, rather than the parameters of the regression model,[9] in fact, we only need to ensure that the predicted probabilities for any non-zero value of the separated covariate and its marginal effects are extremely large or small. To achieve this, it is important to set a suitably large or small (but not necessarily infinite) parameters of the regression model (say, eight for a dummy variable in a binary probit scale) so as to ensure that the predicted probabilities attain their extreme values. This is especially important for an ordered probit model as a very large or small the regression model parameters will also change the cut-off point, and the predicted probabilities for *other* non-separated variables will deviate from their actual values as a consequence.

Therefore, we consider a normal prior with scale 10 for parameters of our ordered probit regression model, which permits the parameter of the separated dummy to reach ± 8 (refer to the red curve in Figure 1.1), and the corresponding predicted probabilities to be sufficiently large or small. A further advantage of this prior is that it has a very small influence on the posterior parameters of the non-separated covariates. As noted by Koop et al. (2007), the normal prior with scale 10 can generally be regarded as a non-informative prior on linear regression parameters. Moreover, our application might unfortunately include many separated covariates. To be on safe ground, we adopt the Cauchy prior with scale 10, which has fatter tails than the $\mathcal{N}(0, 10)$ prior (see the red curve in Figure 1.1), to enhance the likelihood that the predicted probabilities for any non-zero value of the separated covariate and its marginal effects successfully reach their

[9] Except that one is interested in the log-odds ratio of a logit model, e.g. in epidemiological studies.

extremes. It is worthwhile to mention that alternative prior distributions, for example the student's t distribution with small degrees of freedom and a large scale, might function similarly. However, we find that the Cauchy prior with scale 10 behaves relatively better in terms of stability. In the next section, we provide supportive evidence on the choice of the Cauchy prior by using cross-validation[10] to examine models with different prior parameters and different link functions.

1.2.5 SAMPLING FROM POSTERIOR DENSITIES

Before starting the MCMC iterations, we need to initialise the values of Y_i as we do not observe the outcome $Y_i^{(0)}$ in reality. We randomly generate $Y_i^{(0)} \in \{1, 2, 3, 4\}$ in ways such that they are consistent with the actual numbers of aggregate submissions meeting each quality standard (i.e. $T_{js} = \sum_{i=1}^{n_j} I(Y_i = s)$, for $s = 1, 2, 3, 4$). This step can be thought of as an additional initialisation stage of the MCMC scheme. Hence, to sample from the posterior, the Gibbs cycle operates on the variables (β, c, Y) as follows.

Sampling Model Parameters $\theta = (\beta, c)'$ Given y

Given the current state of y, which is obtained from the exchange of probabilities, the target distribution of the parameters $\theta = (\beta, c)'$ after integrating out the latent y^* is:

$$f(\theta|y) \propto \left(\prod_{j}^{J}\prod_{i}^{n_j} f(y_i|\theta)\right)p(\theta|\theta_0) \tag{1.2.8}$$

where $f(y_i|\theta) = \Phi(c_{y_i} - \beta'\mathbf{D}_i) - \Phi(c_{y_i-1} - \beta'\mathbf{D}_i)$ and $p(\theta|\theta_0)$ is the Cauchy prior for parameters. θ_0 consists of the mean and the scale hyperparameters. The mean hyperparameters is fixed at zero, and the scale hyperparameters is 10 for each β and 100 for each cut-off point.

A Metropolis-Hastings sampler with the Laplace approximation is applied to facilitate the sampling (a similar implementation can be seen from Pitt et al. (2006)). We use a multivariate t-distribution $\mathcal{T}_v(\hat{\theta}, V)$ as our proposal distribution in which the mean $\hat{\theta}$ is a

[10]Similarly, Gelman et al. (2008) use cross-validation for prior parameters in a binary logit model.

mode of the log-likelihood and the scale V is the negative inverse of the second derivative, which can be obtained by a Newton or quasi-Newton procedure. The degrees of freedom v are generally chosen at around six to ensure that the proposal density dominates the target in the tails. Then the acceptance probability of candidate $\theta' \sim \mathcal{T}_v(\hat{\theta}, V)$ is:

$$\alpha_\theta = \min\left\{1, \frac{f(\theta'|x)\mathcal{T}_v(\theta|\hat{\theta}, V)}{f(\theta|x)\mathcal{T}_v(\theta'|\hat{\theta}, V)}\right\} \tag{1.2.9}$$

Hence, with probability α_θ, we take $\theta = \theta'$ and otherwise θ stays at current values.

In practice, we utilise the block MCMC to sample β and c separately, which has been found to have a better acceptance rate than that of the case in which they are sampled together. The cut-off points $c^{(s)} = (c_2^{(s)}, c_3^{(s)})'$ are sampled as in a single block. For β, we randomly permutate their order and randomly stratify them into blocks, each of which contains two to six elements, for every iteration. Such block sampling allows us to attain a higher acceptance rate in the Metropolis-Hasting step, as well as allows correlations between the parameters.

Sampling Censored Outcome Variable Y Given θ

Now we specifically illustrate the steps to sample the individual outcome Y using the approach described in Section 1.2.2. According to the current state of $\theta = (\beta, c)'$, two indices i and l are randomly picked from the same institution j. We subsequently compute two configurations:

$$\Pr(Y_i = y_i, Y_l = y_l|\theta) \text{ and } \Pr(Y_i = y_l, Y_l = y_i|\theta) \tag{1.2.10}$$

which is then normalised to compute the probability of exchanging values:

$$\mathcal{P}_{il} := \frac{\Pr(Y_i = y_l, Y_l = y_i|\theta)}{\Pr(Y_i = y_i, Y_l = y_l|\theta) + \Pr(Y_i = y_l, Y_l = y_i|\theta)} \tag{1.2.11}$$

Therefore, y_i and y_l swap values with a probability $\mathcal{P}_{il}$, and both stay at their current values otherwise. In each of the Gibbs sampling iterations, we perform the above steps to exchange the probabilities for every pair of the observations until all have been swept.

1.3 Verification of Algorithm

The objective of this section is to verify our proposed algorithm from different aspects. The first verification is based on an artificial data that is designed to mimic the real data structure. It demonstrates that our model works properly, and the estimated model parameters (β and c) are very close to their benchmark values. Table A.3 and Figure A.1 in the Appendix provide more details of this simulation.

The remaining three verifications we present here are respectively used to 1) verify the rationale for the probability exchange algorithm using true data and simulation data; 2) consider the case in which journal dummies have only a few values of one; and 3) choose proper models and prior hyperparameters by cross-validation.

1.3.1 Rationale for Exchange of Probabilities

Predict the Actual GPA Scores

First we show that the exchange of probabilities algorithm can effectively recover the underlying unobservable individual outcomes in real-life applications, by comparing the observed actual GPA scores of each Economics department with our predicted values (Figure 1.2). The technical details for the calculation of actual GPAs, predicted GPAs, and interval estimates are presented in Appendix A.3.

The sub-plot (1) reveals that the predicted GPA scores remain reasonably close to the true GPA scores along the 45 degree reference line. The two scores are highly correlated as reflected by the correlation coefficient of 0.958. We then fit a regression line of the true scores on the predicted scores with an intercept. The R^2 coefficient is 0.917, and the estimated slope is 1.152 (se=0.058), reflecting the fact that the predicted GPA scores are fairly close to the actual ones. The 95% confidence intervals of the fitted regression line, which almost cover the 45 degree reference line, reassure us that the predicted GPAs and the true GPAs are very similar on average. Sub-plots (2)-(4) of 1.2 for the GPA proportions of 4's, 3's and 2's tell pretty much the same story. In the empirical results in Section 1.4, we present further investigation into the institutional effects.

The Role of Aggregate Information

The basic idea behind the exchange of probabilities lies in the aggregate information available in the REF data. We intuitively believe that this step is likely to induce a more accurate retrieval of the outcomes provided that more aggregate information is available. Thus it is interesting to understand how strong the aggregate information should be in order for us to obtain an accurate recovery and credible estimation results. To illustrate this, suppose that we have aggregate information of about 2,600 individual submissions from 28 institutions, we want to ask: given the information on these 28 institutions, how will our method behave (better or worse) if there are more or fewer individual submissions?

To answer this question, we perform the first simulation whilst relying on an artificial dataset generated by following the data structure of the 2014 REF Economics and Econometrics sub-panel. To be more precise, we define 28 artificial institutions, each of which has the same proportion of submissions that are considered to be of quality 4,* 3*, 2* and 1* as they have in the real data. The values of the LHS outcomes are then simulated to be consistent with the actual numbers of the aggregate proportions for each of the 28 economics departments, as informed by the REF data. We employ four independent covariates in the process of data generation: x_1 and x_2 are Gaussian covariates with a unit variance, and x_3 and x_4 are Bernoulli variables with a probability 0.3 of being one. The actual parameters are 0.8, -0.8, 1.4 and -1.4, respectively. To assess the performance of the algorithm with different sample sizes, we generate artificial datasets with sample sizes 500, 1,500, 2,600, 3,500, 4,500 and 5,000, each repeated four times.

The so-called Hamming loss is utilised to evaluate the extent to which our method for exchanging probabilities can restore the categorical outcome variables. The Hamming loss statistic can be written as:

$$L_{\text{Hamming}}(Y, \hat{Y}) := \frac{1}{n}\sum_{i=1}^{n} I(Y_i \neq \hat{Y}_i) \tag{1.3.1}$$

where Y denotes the actual outcome and $\hat{Y}$ denotes the predicted outcome. In essence, the Hamming loss measures whether the outcome Y can be exactly retrieved exactly. It is suitable for dummy or nominal outcomes that can be specified as true or false.

Figure 1.3 presents the MCMC trace plots of the Hamming loss for selected sample sizes 500, 1,500, 2,600 and 5,000. The Hamming loss has randomly assigned initial values of about 0.75, which can be viewed as the benchmark values. The pattern of the two trace plots reveals that, for all sample sizes, the Hamming loss converges from its benchmark values to the steady state in fewer than 1,000 iterations, suggesting that our algorithm, built on the Metropolis-Hastings sampler, is able to efficiently retrieve the outcome variables.

The Hamming loss has the largest number (about 0.07) when the sample size is 500, which implies that approximately 93 percent of the outcome values are exactly recovered. When there are 2,600 submissions, the same number of submissions as in the real data, the Hamming loss (0.20) suggests that roughly 80 percent of the total outcomes are successfully predicted. Reassuringly, this is a notably high recovery rate relative to the benchmark Hamming loss score (i.e., 0.75 for the outcome with 4 categories). As the Hamming loss only considers exact recovery cases, we see that it clearly gets larger as the sample size increases.

1.3.2 Journal Dummies with Only Few Values of One

Another challenge to our method arises when many dummy variables, of which only very few have the value one, are included (i.e., journal dummy variables have only a few submissions). Consider a toy example with 22 observations and two journal dummy variables: Journal A and Journal B. Assume that Journal A has two submissions to the REF and Journal B has 20. Also assume that each submission has a 50 percent chance of being classified as 4* and a 50 percent chance to be classified as 1*. Therefore, there are two extreme cases for both journals—all submissions are judged as 4* and all submissions are judged as 1*. In the former (latter) case, the coefficients on the journal dummies will be at their largest (smallest) values. All submissions from Journal A have a $0.50^2 = 0.25$ probability of having the quality standard 4* or 1*, while all submissions from Journal B have only a

$0.50^{20} = 9.5 \times 10^{-6}\%$ probability of being classified as 4* or 1*. This suggests that Journal A, which has a smaller number of submissions, has an exponentially higher probability to attain the extreme value, and thus its sampled coefficient will be much wavier than that of Journal B.

Furthermore, if a large fraction of submissions is from journals with only a few submissions, the fluctuations of the sampled coefficients will render waving sampled cut-off points accordingly. Additionally, the sampled coefficients on other journal dummy variables will deviate according to the change of the cut-off points.

We now exploit an artificial dataset to illustrate the aforementioned points. To generate LHS outcome variables, we implement the same procedure as in simulation one and in the previous subsection. The resulting sample size is 2,600, indicating 2,600 submissions. We define six artificial, mutually exclusive journal dummy variables, each of which has 100 submissions. We then generate a series of mutually exclusive dummies for small submissions. A total of eight designs are examined:

Design 1: There is no journals with less submissions.

Design 2: 5% of the total submissions are from journals with 10 submissions.

Design 3: 10% of the total submissions are from journals with 10 submissions.

Design 4: 15% of the total submissions are from journals with 10 submissions.

Design 5: 20% of the total submissions are from journals with 10 submissions.

Design 6: 20% of the total submissions are from journals with 20 submissions.

Design 7: 20% of the total submissions are from journals with 40 submissions.

Design 8: 20% of the total submissions are from journals with 80 submissions.

Table 1.2 reports the posterior means and the standard deviations for the aforementioned designs. In general, the posterior means of all these designs, presented in columns (4)-(11), are close to the benchmark estimates by the maximum likelihood or MCMC, given that the true Y is known (columns (2)-(3)). In columns (4)-(8) we observe that, as the proportion of journals with 10 submissions rises from zero percent to 20 percent, the posterior means of the parameters of journals with 100 submissions differ more significantly from their benchmark values. Columns (8)-(11) correspond to the results of designs 5-8.

The coefficient estimates suggest that, when conditioning on 20 percent of total submissions and when the number of submissions in each journal increases (from 10 to 80), the posterior means become much closer to the benchmark values. However, there is no significant difference in the posterior standard deviations of β. Regarding the cut-off points, we observe a similar pattern in the results. In addition, the posterior standard deviations of the cut-off points go larger from design 4 to design 8, but turn smaller from design 8 to design 11.

Crucially for the analysis, the effects of main interest pertain to the predicted probabilities for each journal dummy. Interestingly, the posterior probabilities, reported in Table 1.3, show that there is almost no significant difference in terms of the probabilities of journals with 100 submissions, for all eight designs. We do not find a sound analytical justification for this. One possible explanation is that although the journals with small submissions fluctuate in both β and c, the changes in β and c are in the same direction and eventually offset each other in the equation that calculates the probabilities (Equation 1.2.2). In summary, there are two implications for the inclusion of journal dummies with a small number of submissions. First, when a large proportion of total submissions comes from journals with only a few submissions (e.g. 10 submissions), the sampled cut-off points become waving and the posterior means of the parameters of other journal dummies differ significantly from their benchmark values. Second, it seems that this does not lead to any significant difference in the probabilities of these journal dummies because the fluctuations can cancel each other out.

1.3.3 Cross-validation

The last simulation is to explore a suitable prior distribution and a model link function (i.e., a probit or logit) for our empirical applications. We use five-fold cross-validation to evaluate the model fitness under different specifications. The idea of the cross-validation originates from an interesting line of research on machine learning, which has commonly been used to evaluate model predictions. Here we implement a five-fold cross-validation—the same number of folds used in Gelman et al. (2008)—for a binary logit model. In particular, we randomly split the data into five sub-samples of equal size. Four out of the five sub-samples are used as training data, and the remaining sub-sample as validation data. In the ordered probit model we use the parameters estimated by the training data to predict the outcome in the validation data. We focus on two statistics to examine the quality of our predictions. The first is the Brier score loss (Brier, 1950), defined as the mean square error from models with dichotomous outcomes (e.g. Bernhardt et al. (2016) applied Brier score loss for the ordered probit model):

$$B_{i,\text{loss}} \coloneqq \sum_{s=1}^{4} (I(Y_i = s) - \Pr(Y_i = s))^2 \tag{1.3.2}$$

The second statistic refers to the log loss, which is defined on the estimated probabilities rather than the discrete predictions. It can be written as a negative log-likelihood:

$$L_{i,\text{loss}} \coloneqq -\sum_{s=1}^{4} I(Y_i = s) \log \Pr(Y_i = s) \tag{1.3.3}$$

The process above is rotated for all five sub-samples so that each observation will be used in the validation. We repeat the cross-validation procedure for different combinations of the link functions, scales and degrees of freedom of the student's t prior.

It is important to note that the focus of this subsection is not on the recovery of outcomes. Therefore, we fix the outcome variables Y at their initial assignments, following Hole (2015). Specifically, the values of Y are simulated according to the predicted probabilities for each economics journal that meets the REF criteria as listed in Table A1 in Hole (2015). We then define two different sets of journal dummy variables (65 dummies and 37 dummies) which coincide with the specifications explored in the next section.

The average Brier score loss and average log loss are plotted in Figure 1.4-1.5. For the probit model, both figures imply that, when the scale is greater than about nine, the student's priors (no matter the scale parameter) have similarly small Brier score and log losses under the probit specification. We find that the Cauchy prior (labelled as the t distribution with degrees of freedom one) seems to have the lowest Brier score and log losses for all scale parameters. In general, the probit model appears to show lower loss statistics than the logit model does. From the four subplots in Figure 1.4-1.5, it is still difficult to say which prior specification is predominantly good. Therefore, it is safe to conclude that the probit model fits our empirical work better than the logit model, and that the Cauchy prior with a scale around 10 under the ordered probit set-up is a consistently better choice.

1.4 Empirical Results and Analysis

1.4.1 Background and Data

We draw large scale data from the 2014 Research Excellence Framework (REF) to facilitate the empirical analysis of journal perceptions.[11] The REF is a framework for assessing research quality and impact in UK Higher Education Institutions (HEIs), last conducted by the UK government in 2014.[12] It aims to support a dynamic and competitive UK research environment and to provide information for academic and public use. The quality of each submitted research output (e.g. journal article, working paper, monograph, or book chapter) published between January 2008 and December 2013 was assessed by peer reviewers from 36 expert sub-panels in different research areas. The outcome was in the form of a quality profile, which includes one of the following classifications: "world leading" (4*), "internationally excellent" (3*), "recognised internationally" (2*), "recognised nationally" (1*), and "unclassified".[13] Our algorithm effectively merges the last two categories as they both indicate poor quality of the work.

Table 1.1 presents descriptive statistics for 2,600 submissions to the REF Economics and Econometrics sub-panel at the level of the institution. In the 2014 exercise, the largest submission to the Economics and Econometrics unit came from the University of Oxford, which submitted 242 research outputs, and the smallest came from the University of East

[11]All data is publicly available from the REF official website: http://results.ref.ac.uk/.

[12]The REF replaced the previous Research Assessment Exercise (RAE) in 2008. It was carried out jointly by four authorities: the Higher Education Funding Council for England (HEFCE), the Scottish Funding Council (SFC), the Higher Education Funding Council for Wales (HEFCW) and the Department for Employment and Learning, Northern Ireland (DEL).

[13]The work in the unclassified category has the quality that falls below 1* or does not meet the published standard for the purpose of this assessment.

Anglia, which submitted only 49. The overall distribution of the average quality increases from 3.65 percent in 1*, to 20.19 percent in 2*, to 48.65 percent in 3*, and then decreases to 27.5 percent in 4*. This suggests that a large number of research outputs in economics are perceived to be of high quality by members and assessors.

Table A.1 summarises, at the journal level, the number of submissions for all economics outputs to the Economics and Econometrics sub-panel and to the Business and Management Studies sub-panel, presented respectively in column (2) and (3). In this paper, we primarily focus on the perception of economics journals by the Economics and Econometrics sub-panel. We look at the Business and Management Studies sub-panel purely for comparison purposes. Based on the Panel Criteria and Working Methods for REF 2014, the REF exercise operated a few calibration exercises to ensure the consistency of assessment across the sub-panels, including cross-referral process, introduction of a member sitting in both sub-panels, and justifications by sub-panel chairs. Moreover, Pidd and Broadbent (2015) disclosed more details about the Business and Management Studies sub-panel[14]. In reality, the Business and Management Studies sub-panel has passed on more than 1300 economics research outputs to the Economics and Econometrics sub-panel for the cross-referral. Subsequently, it assigned final grades to these economics submissions according to the recommended scores they received from the Economics and Econometrics sub-panel.

However, publicly available information on the REF 2014 did not explicitly indicates that which outputs should be considered as "economics" outputs, nor did it clearly reveal to the public that how the "economics" outputs had been defined by the Business and Management studies sub-panel. Column 2 of Table A.1 lists the 3,055 outputs from the Business and Management Studies sub-panel being published in the same journals that outputs from the Economics and Econometrics sub-panel were also published. We cannot separate out the cross-referred outputs from these 3,055 outputs. Therefore, outputs published at economics journals from the Business and Management Studies sub-panel might

[14] Pidd and Broadbent (2015) is a paper written by the chair and the deputy chair of the Business and Management Studies sub-panel for the REF 2014 exercise, on behalf of this sub-panel.

contain both cross-referred and independent-referred outputs.

Besides, the first column of Table A.1 shows that the 2014 SJR ranking[15] for each economics journal, which serves two important roles. Firstly, it provides useful benchmarking information on the journal ranking for comparison purposes. Secondly, in later analysis, we will aggregate journals with small submissions into several groups according to their positions in the SJR ranking.

1.4.2 Results: Economics and Econometrics Sub-panel

First we present results for journals with at least 20 submissions to avoid any potential bias caused by journals with small submissions, as discussed in Section 1.3.2. We then incorporate dummy variables for journals with only 10 to 20 submissions, and reconcile the two results.

Journals with more than 20 submissions

We start with a description of the estimated results for journals with more than 20 submissions, as summarised in Table 1.4-1.5 and Figure 1.6. In Table 1.5 we report the number of submissions to each journal (column (1)), the corresponding SJR ranking in 2014 for each journal (column (2)), the respective probability of being ranked as 4*/3*/2*/1* (columns (3)-(10)), and the average score by weighting the four predictive rankings (columns (11)-(12)). Unsurprisingly, the five top-ranked journals are AER, Econometrica, REStud, QJE, and JME.[16] Within the top five, the AER (GPA score 3.972), Econometrica (GPA score 3.971) and REStud (GPA score 3.9721) are clearly distinguishable from the remaining two, with more than a 90 percent chance of making it into rank 4* but a zero probability of being ranked as 2* or 1*. Reassuringly, the regression model parameters (β and c) and the associated predicted probabilities exhibit narrow MCMC standard deviations.

[15] The SJR (SCImago Journal Rank), which is a measure of the scientific influence of journals that account, ranks scientific journals of 27 main subject areas and 313 specific subject categories, based on the citation data drawn from over 21,500 titles and from more than 5,000 international publishers. More details of the SJR indicator can be found at SCImago (2007).

[16] To interpret results neatly, some of the generally recognisable economics journals are abbreviated. For example, the American Economic Review is abbreviated as AER, and the Journal of Political Economy as JPE. The full list of abbreviations can be found in Table A.2 of the Appendix.

The journals ranked sixth through to tenth are JEEA, JPE, Journal of Econometrics, JDE, and JIE, with the GPA socre ranging from 3.500 for the JIE to 3.784 for the JEEA. Contrary to expectations, the JPE (GPA score 3.716) appears to have a relatively low rank. One possible explanation lies in the fact that this journal has the smallest number of submissions (22) in this model specification, and thus may not be very representative. A similar case is observed for the Oxford Bulletin of Economics and Statistics, whose number of submission is only 24, resulting in a wide MCMC standard deviation. The EJ is number 11, the JPubE number 12, the JAE number 13, the REStat number 14 and the JET number 15. The GPA score ranges from 3.045 for the JET to 3.339 for the EJ. Among the top-16 to 20, there is one finance journal (JMCB) and four economics journals (IER, Games and Economic Behaviour, EER, and Economic Theory). The weighted mean quality ratings are between 2.797 and 2.991.

From the 20th-ranked journal on down, the Economics Letters is ranked the lowest, perhaps due to its nature of being a supplement to the specialist literature by providing quick dissemination and accessibility (as described on its official webpage at Elsevier[17]).

Most prominently, the estimates indicate that "other journals" with higher ranks according to the SJR are also ranked unambiguously higher in our model. With regard to the three categories at the bottom, we see that unpublished papers and other types of unpublished work (GPA score 3.409) have a rank between the JIE and the EJ, while book chapters (GPA score 3.219) are ranked between the JAE and the REStat. Other articles (GPA score 1.630) are perceived the least favourably. In addition, it is important to note that the overall pattern of the rankings generated by our algorithm is generally consistent with that of the SJR 2014.

Journals with 10-20 submissions

Turning to journals with 10-20 submissions, some interesting results emerge from Table 1.6. First, several journals introduced in recent years dominate at the top. For instance, the Quantitative Economics, which began publishing in 2009, is ranked first in this list,

[17] The website of the Economics Letters is at: http://www.journals.elsevier.com/economics-letters.

with the highest probability of being perceived to be of 4* quality (0.909) and the highest GPA score at 3.909. Another successful newcomer is the American Economic journal series, which also started in 2009, and occupies three out of the five top positions. Specifically, the AEJ: Economic Policy, Microeconomics, and Applied Economics have, respectively, 87.6 percent, 59.0 percent, and 51.1 percent chances of being considered to be of the highest standard. Second, differences in journal quality perceptions appear to exist across different fields, and our algorithm seems to rank the qualitative and applied journals more highly. A notable example is given by the journals from economic history, such as the Journal of Economic History and the Explorations in Economic History, which are ranked at second and seventh, respectively. Among those more theoretical journals, however, the rankings decline sharply (e.g. Journal of Mathematical Economics, AEJ: Macroeconomics, and Journal of Business and Economic Statistics).

The second panel of the table displays the results for journals with more than 20 submissions after adding journals with 10-20 submissions to the estimation. The posterior predicted probabilities in the second panel of Table 1.6 are very similar to the ones in Table 1.4, which echoes our simulation evidence (Section 1.3.2) that journals with fewer submissions could have a large impact on the posterior of other parameters of the regression but have a limited impact on the resulting predicted probabilities. Compared with the previous results in Table 1.4, a great deal of the journals are consistently ranked after accounting for observations with 10-20 submissions. The stability of the ranking is apparent at the very top of the ranking—the top five journals (AER, Econometrica, REStud, QJE, and JME) are exactly the same for the two samples. We note an erratic behaviour of the ranking for a small number of journals, but the fluctuations in journal ranks are mostly within adjacent categories. For instance, the JPE moves upward by just one category from number seven to number six, while the JEEA moves downward by two categories from number six to number eight.

Institutional Fixed Effects

Furthermore, we re-investigate Figure 1.2 to explore whether there are institutional effects on the REF 2014 perception. Encouragingly, we find that the predicted and the true GPA scores are almost identical for institutions with mid-ranged GPA scores, such as Edinburgh, Essex, Glasgow, Leicester, and Manchester. Interestingly, institutions with higher GPA scores perform better than the model predicts (i.e., UCL, Cambridge, Warwick, and Oxford), whereas institutions with lower GPA scores perform worse than the model predicts (i.e., Brunel, Kent, and Aberdeen).

The evidence above implies the existence of institutional fixed effects on journal quality perception by the REF. In other words, given a particular type of submission or publication at the same journal, outputs from recognisably prestigious Economics departments such as those of UCL, Cambridge, Warwick, and Oxford would have a greater chance of being rated highly by the REF panel. Conversely, submissions from less prestigious Economics departments at institutions such as Brunel, Kent, and Aberdeen would receive a poor perception by the REF. For instance, a working paper from Oxford is more likely to attain a higher REF criteria than one from Kent; or an EJ publication from UCL is more likely to be rated as having good quality than the same publication from Brunel.

1.4.3 Results: Economics Journals in Business and Management Studies Sub-panel

Next we examine how our model predicts the perception of quality for economics journals by the REF Business and Management Studies sub-panel. We focus on those typical economics journals with more than 10 submissions. According to the results in Table 1.7 and Figure 1.7, we see that as the proportion of economics journals with 10-20 submissions is relatively small (220 out of a total of 12,171 submissions from 98 institutions), the MCMC sampled cut-off points are persistent and strongly mixing (Figure A.5 in Appendix), which reconciles our demonstration in Section 1.3.2 that a very small subset of the sample from journals with small submissions will not make cut-off points further waving.

The REStud has the highest ranking, followed by the JIE and the JME (with respective weighted average quality ratings being 3.964, 3.943, and 3.934). The top general interest economics journals (i.e. AER and REStud) still maintain a high perception in the Business and Management Studies sub-panel. Besides, a comparison of the results of economics journals perceived by the Business and Management Studies sub-panel (Table 1.7) with similar results provided by the Economics and Econometrics sub-panel (Table 1.4 and Table 1.6) suggests that the Business and Management Studies assessment unit holds a better perception concerning the quality of economics journals. For example, the JET is estimated to have the probability of 0.156 being ranked as 4* in the Economics and Econometrics sub-panel, while the estimated probability is 0.930 here. Similarly, the weighted average quality rating for Games and Economic Behaviour was 2.882 before but became 3.863 now. Significant upward movements are also observed for the EJ, Journal of Health Economics, Journal of Economic Dynamics and Control, among others. Taken as a whole, this indicates that the perception of economics journal quality is significantly higher in the Business and Management Studies sub-panels.

Since the REF 2014 took a few measures to ensure the consistency of judgement across various sub-panels as mentioned in Section 1.4.1, one possible explanation of the difference in the results is that the overall quality of the economics submissions produced in the Business and Management sub-panel is generally higher than that in the Economics and Econometrics sub-panel.

1.5 Conclusion

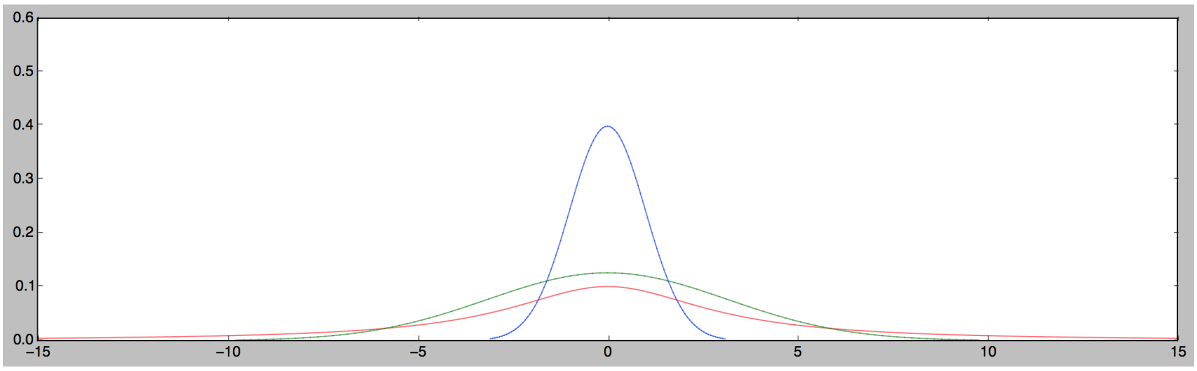
In this paper we aim to investigate how the REF 2014 perceives journals, with a particular interest in the economics journals. We began by highlighting the problem that we cannot observe the outcome that how each individual output admitted by the REF 2014 exercise is perceived, but we do know the aggregate counts/proportions of how many submissions from each institution meet the REF 2014 criteria. We pursue this idea by working within the framework of ordered probit regression, and propose a new approach to sample the unobservable outcomes. A common concern with a probit or logit model relates to the separation problem. We overcome this potential issue by incorporating a diffused prior to our proposed model—the Cauchy distribution with scale 10. We develop an efficient MCMC scheme to sample parameters of the ordinal response regression model, as well as the outcome variable. We verify the algorithm on various simulations with an artificial dataset that mimics the data structure of the REF 2014.

Our empirical evidence mainly focuses on the economics journals' submissions and other types of submissions to the Economics and Econometrics sub-panel. Moreover, we also checked the Business and Management Studies sub-panel to investigate whether submissions from same economics journals would have been treated differently in another REF sub-panels.

In addition, the methodology of this paper can easily be extended to any generalised linear regression models, and could serve as a solution to estimate the effect in the case that the outcome is only available as aggregate information for sub-groups.

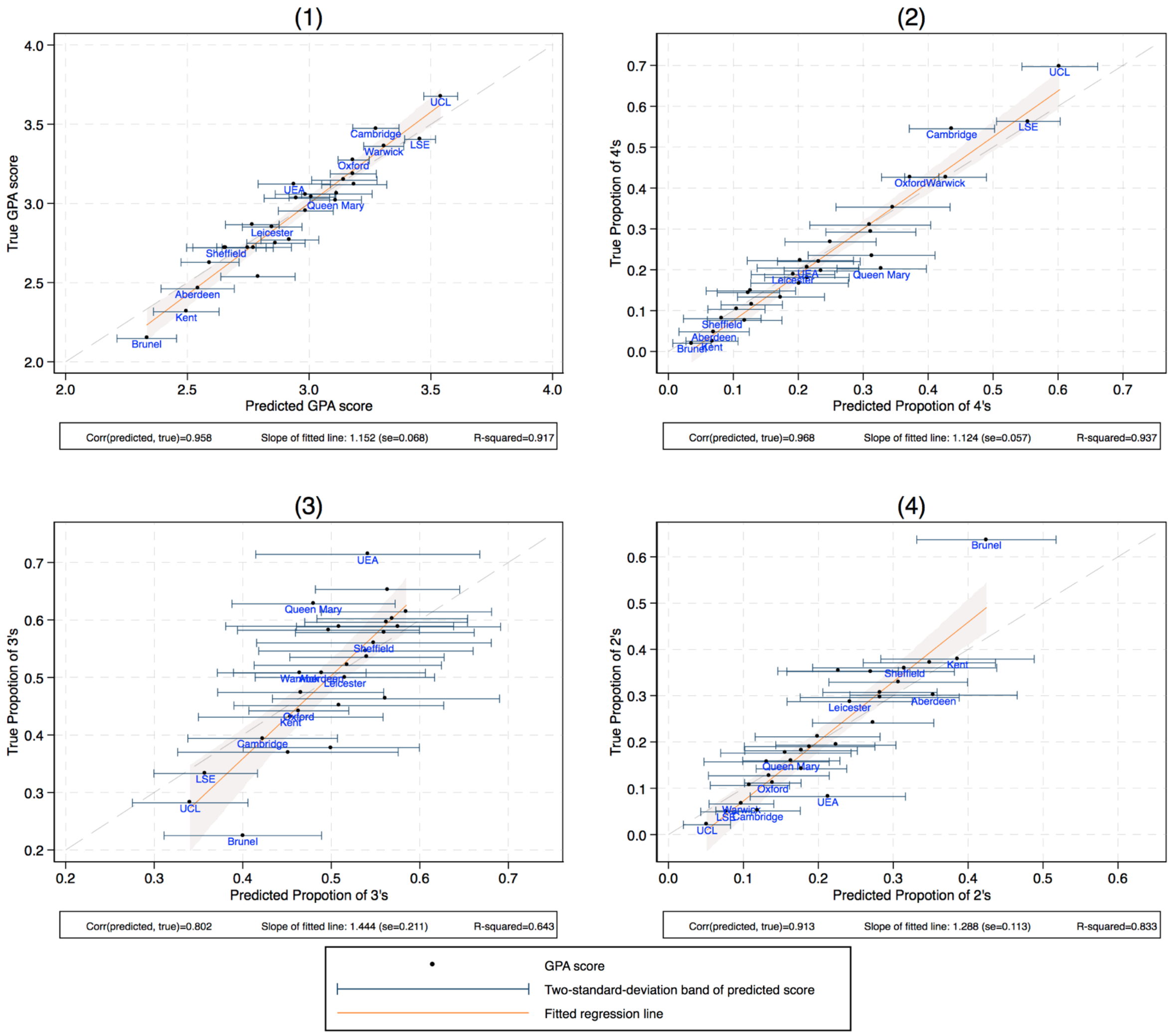
Figures and Tables

Figure 1.1: DENSITIES OF THE STANDARD NORMAL, $\mathcal{N}(0, 10^2)$ AND $Cauchy(0, 10)$



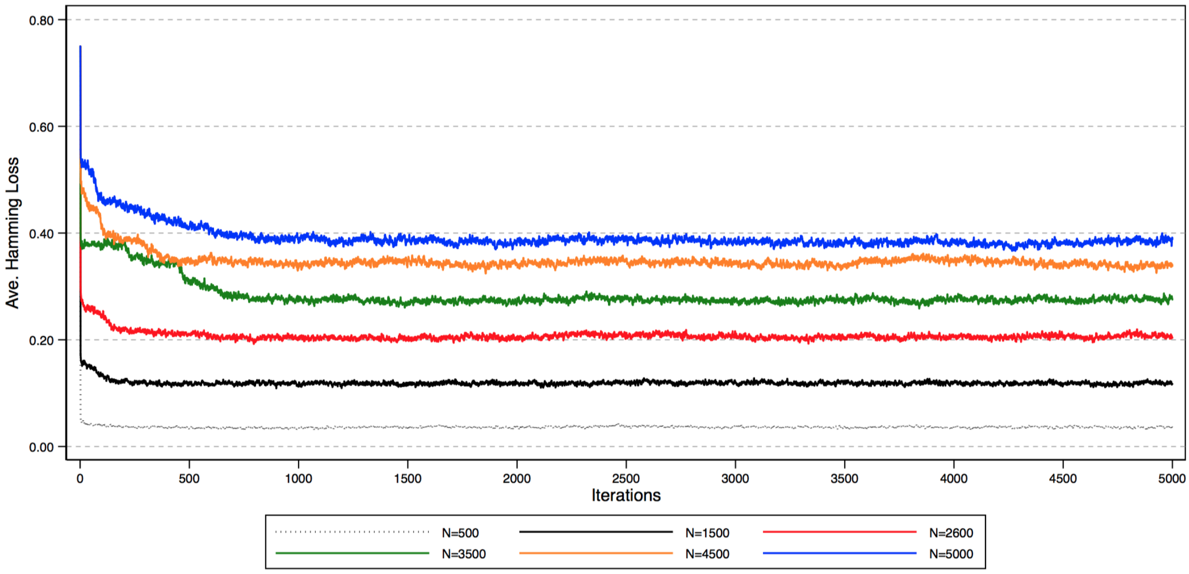
NOTE.—The *blue line* refers to the density of the standard normal distribution (Probit link). The *green line* refers to he density of zero mean normal with scale 10. The *Red line* refers to the zero mean Cauchy distribution with scale 10.

Figure 1.2: Predicted and True GPA Score of Each Institution by REF Economics and Econometrics Panel



Note.—The results are obtained by 20,000 iterations after 3000 burn-in periods of the proposed MCMC algorithm base on data from the REF 2014 Economics and Econometrics Assessment Unit. Sub-plot (1) lists the number of submission of each journal. Sub-plots (2)-(4) reflect the proportions 4's, 3's and 2's of the GPA score.

Figure 1.3: Trace Plots of the Average Hamming Loss by Different Sample Sizes – Simulation Data



Note.—Each curve represents the Hamming loss by averaging over four instances of independent chains, each of which is based on a randomly generated sample according to the data generating process defined in the text.

Figure 1.4: Average Brier Score Loss and Log Loss for Different Scales and Degrees of Freedom of the Student-T Prior and for Different Link Functions of the Ordered Response Model – Five-fold Cross-Validation (65 covariates)

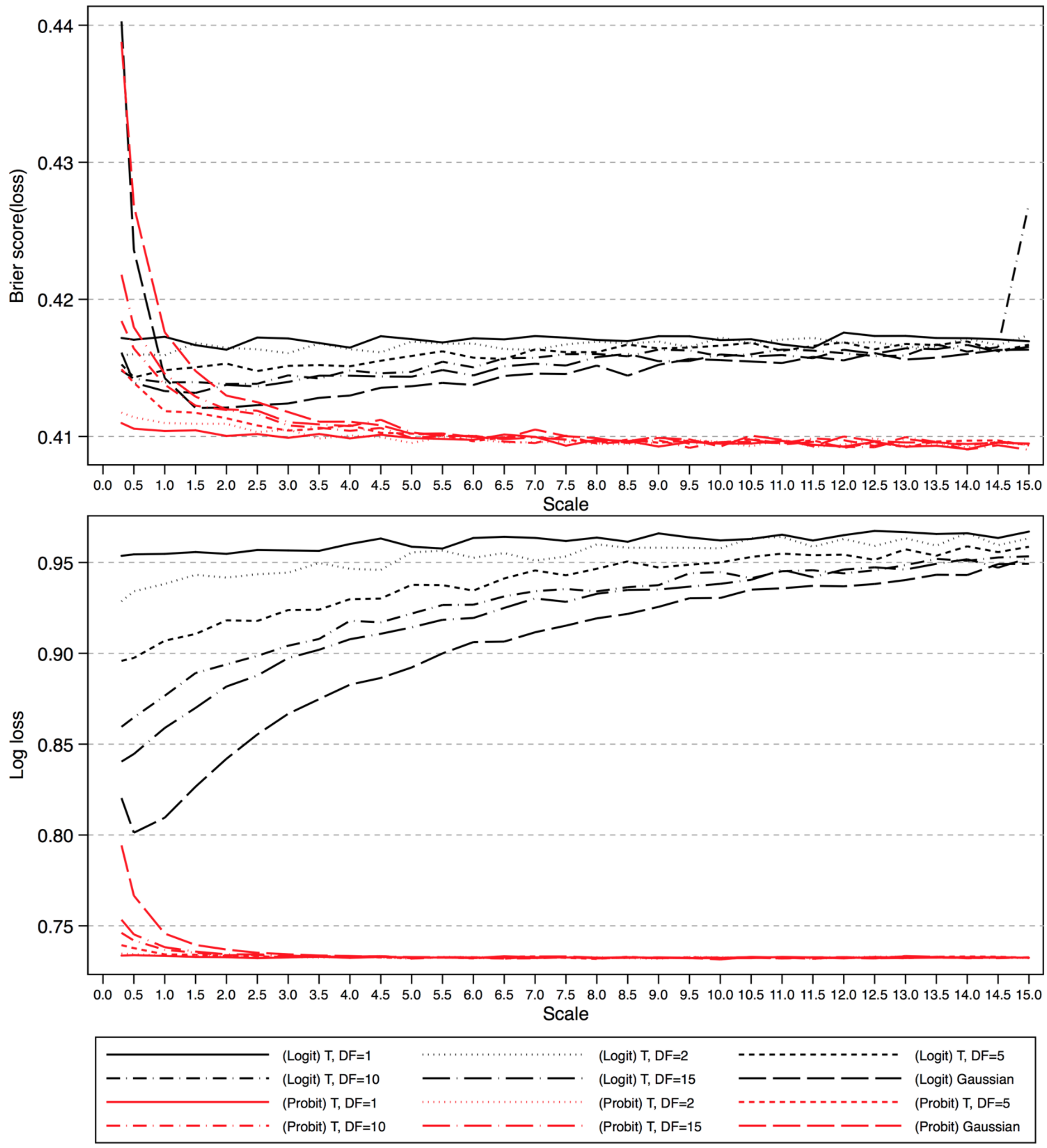


Figure 1.5: Average Brier Score Loss and Log Loss for Different Scales and Degrees of Freedom of the Student-T Prior and for Different Link Functions of the Ordered Response Model – Five-fold Cross-Validation (37 covariates)

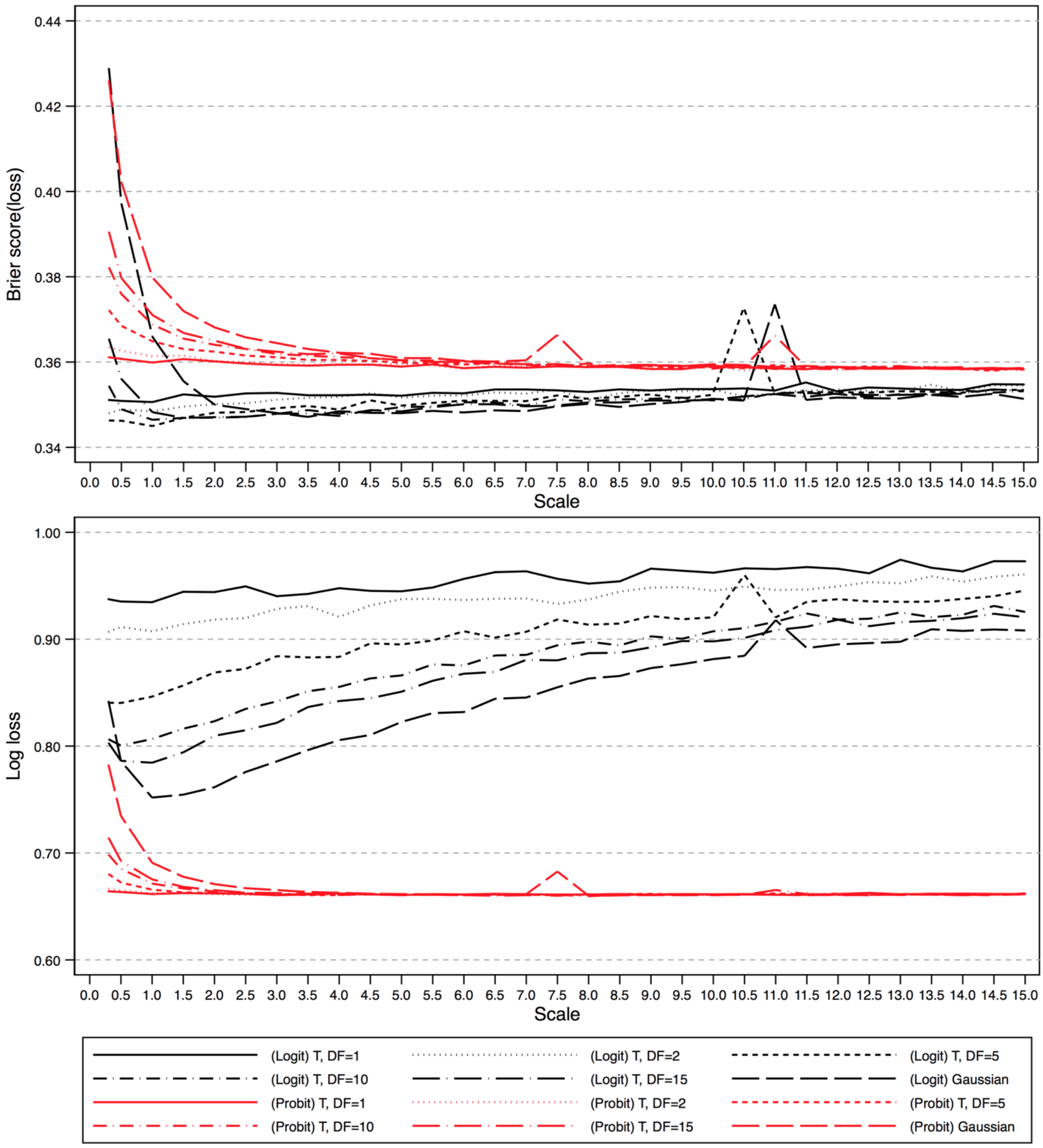


Table 1.1: Proportion of Submissions to REF Economics and Econometrics Panel Meeting REF Quality Standards 1, 2, 3, 4 by 28 Institutions

		Percentage of the submission meeting the standard for:			
Institution	**Number of submissions**	**1***	**2**	**3**	**4**
	(1)	(2)	(3)	(4)	(5)
University of East Anglia	49	0	8.2	71.4	20.4
University of Sheffield	50	0	36	56	8
Royal Holloway London	51	3.9	15.7	45.1	35.3
University of St Andrews	51	0	17.6	58.9	23.5
City University London	54	16.7	29.6	37	16.7
University of Sussex	54	3.7	35.2	46.3	14.8
University of Edinburgh	55	1.8	12.7	54.6	30.9
University of Aberdeen	63	14.3	30.1	50.8	4.8
University of Bristol	63	0	19	58.8	22.2
University of Surrey	71	0	21.1	52.1	26.8
University of Birmingham	79	1.3	32.9	58.2	7.6
University of Kent	79	16.5	37.9	43.1	2.5
University of Leicester	80	2.5	28.7	50	18.8
University of Southampton	82	4.9	35.3	37.8	22
University of Exeter	83	9.6	19.3	57.8	13.3
University of Glasgow	83	2.4	18.1	61.4	18.1
Queen Mary London	94	1.1	15.9	62.8	20.2
Birkbeck College London	97	5.2	37.1	47.4	10.3
University of Cambridge	99	1	5.1	39.4	54.5
Brunel University London	102	11.8	63.7	22.5	2
University of York	104	1.9	24.1	59.6	14.4
University of Essex	113	0	10.6	60.2	29.2
University of Manchester	114	4.4	30.7	53.5	11.4
University of Nottingham	127	0.8	14.2	65.3	19.7
University of Warwick	136	0	6.6	50.8	42.6
University College London	142	0	2.1	28.2	69.7
London School of Economics	183	5.5	4.9	33.3	56.3
University of Oxford	242	2.1	11.1	44.2	42.6
Total	2600	3.65	20.19	48.65	27.50

Source.—REF 2014 Economics and Econometrics Assessment Unit Results.

* The unclassified category is merged into the first category.

Table 1.2: ESTIMATED MODEL PARAMETERS BY DIFFERENT PROPORTIONS OF JOURNALS WITH SMALL SUBMISSIONS – SIMULATION DATA

	Benchmarks			Percent of journals with only 10 submissions					Number of submissions in each journal (conditional on 20% of total submissions)		
	True value	MLE (known Y)	MCMC (known Y)	0%	5%	10%	15%	20%	20 submission	40 submissions	80 submissions
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
β_1	2	1.953	1.953	2.147	2.18	2.163	2.26	2.281	2.159	2.196	2.105
		(0.120)	(0.088)	(0.213)	(0.250)	(0.224)	(0.217)	(0.254)	(0.223)	(0.223)	(0.234)
β_2	1	1.014	1.015	0.574	0.588	0.638	0.632	0.666	0.631	0.639	0.603
		(0.107)	(0.079)	(0.197)	(0.215)	(0.213)	(0.185)	(0.225)	(0.190)	(0.184)	(0.207)
β_3	2	2.306	2.307	2.334	2.308	2.373	2.361	2.427	2.4	2.391	2.317
		(0.108)	(0.083)	(0.243)	(0.233)	(0.252)	(0.253)	(0.244)	(0.239)	(0.242)	(0.254)
β_4	1	1.143	1.143	1.227	1.28	1.315	1.387	1.447	1.319	1.19	1.235
		(0.107)	(0.077)	(0.211)	(0.215)	(0.238)	(0.231)	(0.222)	(0.213)	(0.230)	(0.199)
β_5	2	2.059	2.059	2.311	2.335	2.352	2.35	2.443	2.369	2.346	2.292
		(0.124)	(0.090)	(0.206)	(0.211)	(0.232)	(0.236)	(0.236)	(0.212)	(0.214)	(0.222)
β_6	1	1.229	1.231	1.216	1.226	1.201	1.403	1.477	1.371	1.219	1.231
		(0.113)	(0.079)	(0.222)	(0.233)	(0.237)	(0.234)	(0.213)	(0.229)	(0.241)	(0.221)
Cutoff 1	1.868	1.951	1.952	2.041	2.102	2.082	2.158	2.23	2.11	2.098	2.079
		(0.053)	(0.046)	(0.067)	(0.098)	(0.090)	(0.103)	(0.146)	(0.085)	(0.092)	(0.082)
Cutoff 2	1.065	1.143	1.144	1.207	1.237	1.225	1.269	1.301	1.225	1.219	1.208
		(0.048)	(0.035)	(0.048)	(0.065)	(0.059)	(0.067)	(0.086)	(0.054)	(0.058)	(0.053)
Cutoff 3	2.389	2.498	2.499	2.593	2.648	2.652	2.715	2.775	2.647	2.63	2.602
		(0.057)	(0.040)	(0.057)	(0.079)	(0.073)	(0.081)	(0.106)	(0.065)	(0.072)	(0.063)

NOTE.—The results are obtained by 5000 iterations after 1000 burn-in periods of the proposed MCMC algorithm based on artificial data. Columns (1)-(3) represent benchmark values: column (1) displays true values in the data generating process; column (2) and (3) respectively show the maximum likelihood estimates and MCMC estimates given known outcomes Y. Columns (4)-(8) show the results pertaining to different proportions of journals with only 10 submissions. The results throughout columns (9)-(11) are based on simulation data by fixing the proportion of total submissions from journals with 20, 40 and 80 submissions at 20%. MLE estimated standard errors or MCMC posterior standard deviations are reported in parentheses. Model parameters are the parameters of the ordinal regression model (i.e. β and c)

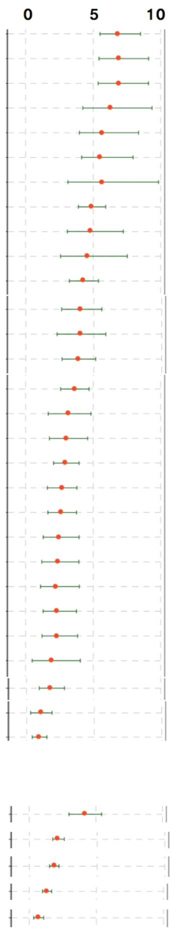
Table 1.3: ESTIMATED MARGINAL EFFECTS BY DIFFERENT PROPORTIONS OF JOURNALS WITH SMALL SUBMISSIONS – SIMULATION DATA

		Percent of journals with only 10 submissions										Number of submissions in each journal (conditional on					
	True	0%		5%		10%		15%		20%		20 submission		40 submissions		80 submissions	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
$Pr(\beta_1 = 4)$	0.349	0.294	(0.028)	0.334	(0.074)	0.318	(0.076)	0.319	(0.074)	0.318	(0.078)	0.313	(0.075)	0.329	(0.078)	0.319	(0.075)
$Pr(\beta_1 = 3)$	0.476	0.497	(0.009)	0.49	(0.026)	0.503	(0.025)	0.51	(0.026)	0.516	(0.028)	0.505	(0.025)	0.495	(0.028)	0.495	(0.023)
$Pr(\beta_1 = 2)$	0.152	0.184	(0.019)	0.159	(0.044)	0.161	(0.048)	0.157	(0.046)	0.153	(0.048)	0.164	(0.047)	0.158	(0.047)	0.167	(0.047)
$Pr(\beta_1 = 1)$	0.023	0.026	(0.005)	0.017	(0.009)	0.017	(0.010)	0.014	(0.009)	0.012	(0.008)	0.017	(0.009)	0.017	(0.010)	0.019	(0.011)
$Pr(\beta_2 = 4)$	0.082	0.07	(0.011)	0.023	(0.011)	0.025	(0.014)	0.024	(0.013)	0.025	(0.015)	0.023	(0.013)	0.026	(0.013)	0.025	(0.015)
$Pr(\beta_2 = 3)$	0.392	0.379	(0.021)	0.237	(0.053)	0.258	(0.061)	0.265	(0.058)	0.276	(0.070)	0.251	(0.060)	0.258	(0.057)	0.247	(0.065)
$Pr(\beta_2 = 2)$	0.367	0.396	(0.016)	0.447	(0.015)	0.451	(0.021)	0.466	(0.024)	0.472	(0.027)	0.455	(0.020)	0.448	(0.020)	0.445	(0.020)
$Pr(\beta_2 = 1)$	0.159	0.156	(0.019)	0.294	(0.061)	0.266	(0.065)	0.245	(0.058)	0.227	(0.066)	0.27	(0.064)	0.267	(0.061)	0.283	(0.071)
$Pr(\beta_3 = 4)$	0.349	0.424	(0.030)	0.402	(0.083)	0.394	(0.087)	0.38	(0.089)	0.397	(0.089)	0.396	(0.083)	0.406	(0.083)	0.401	(0.082)
$Pr(\beta_3 = 3)$	0.476	0.453	(0.016)	0.464	(0.039)	0.475	(0.040)	0.485	(0.042)	0.484	(0.045)	0.475	(0.039)	0.466	(0.040)	0.466	(0.038)
$Pr(\beta_3 = 2)$	0.152	0.112	(0.013)	0.123	(0.040)	0.121	(0.044)	0.125	(0.045)	0.112	(0.042)	0.119	(0.041)	0.117	(0.040)	0.122	(0.041)
$Pr(\beta_3 = 1)$	0.023	0.011	(0.002)	0.011	(0.007)	0.01	(0.007)	0.01	(0.007)	0.007	(0.005)	0.01	(0.006)	0.01	(0.006)	0.011	(0.007)
$Pr(\beta_4 = 4)$	0.082	0.088	(0.013)	0.091	(0.035)	0.094	(0.038)	0.1	(0.038)	0.092	(0.035)	0.093	(0.037)	0.09	(0.040)	0.092	(0.033)
$Pr(\beta_4 = 3)$	0.392	0.411	(0.019)	0.417	(0.049)	0.435	(0.050)	0.455	(0.046)	0.455	(0.049)	0.436	(0.049)	0.42	(0.056)	0.424	(0.047)
$Pr(\beta_4 = 2)$	0.367	0.373	(0.018)	0.378	(0.045)	0.369	(0.051)	0.362	(0.052)	0.37	(0.051)	0.371	(0.050)	0.377	(0.053)	0.375	(0.044)
$Pr(\beta_4 = 1)$	0.159	0.127	(0.016)	0.114	(0.039)	0.102	(0.038)	0.084	(0.032)	0.082	(0.034)	0.1	(0.037)	0.114	(0.045)	0.11	(0.037)
$Pr(\beta_5 = 4)$	0.349	0.331	(0.030)	0.393	(0.073)	0.385	(0.075)	0.384	(0.077)	0.382	(0.080)	0.387	(0.074)	0.393	(0.075)	0.387	(0.072)
$Pr(\beta_5 = 3)$	0.476	0.488	(0.011)	0.469	(0.033)	0.481	(0.035)	0.486	(0.036)	0.492	(0.040)	0.481	(0.035)	0.473	(0.034)	0.474	(0.032)
$Pr(\beta_5 = 2)$	0.152	0.161	(0.018)	0.126	(0.036)	0.124	(0.037)	0.121	(0.039)	0.118	(0.039)	0.123	(0.036)	0.123	(0.037)	0.127	(0.036)
$Pr(\beta_5 = 1)$	0.023	0.02	(0.004)	0.011	(0.006)	0.01	(0.006)	0.009	(0.005)	0.008	(0.005)	0.01	(0.005)	0.011	(0.006)	0.012	(0.006)
$Pr(\beta_6 = 4)$	0.082	0.103	(0.014)	0.089	(0.036)	0.084	(0.037)	0.081	(0.039)	0.088	(0.037)	0.094	(0.042)	0.088	(0.041)	0.089	(0.038)
$Pr(\beta_6 = 3)$	0.392	0.432	(0.017)	0.414	(0.052)	0.419	(0.061)	0.42	(0.066)	0.447	(0.055)	0.433	(0.055)	0.416	(0.057)	0.414	(0.056)
$Pr(\beta_6 = 2)$	0.367	0.355	(0.018)	0.381	(0.047)	0.382	(0.052)	0.39	(0.057)	0.378	(0.055)	0.371	(0.055)	0.38	(0.054)	0.379	(0.050)
$Pr(\beta_6 = 1)$	0.159	0.109	(0.014)	0.117	(0.041)	0.115	(0.048)	0.11	(0.049)	0.088	(0.038)	0.102	(0.042)	0.117	(0.045)	0.118	(0.045)

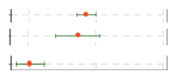
NOTE.—The corresponding marginal effects for Table 1.2.

Table 1.4: Estimated Model Parameters and MCMC Diagnostic Statistics (Journals with More than 20 Submissions to REF Economics and Econometrics Panel)

	No. of submis-sions	MCMC		95%	Batch		Autocorrelation			
		Mean	Std Dev	Asymetric HPD	Std Err	ACRF	Lag 1	Lag 10	Lag 50	Lag 100
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Journals with more than 20 submissions										
American Economic Review	108	6.810	(0.802)		0.033	0.459	0.887	0.589	0.229	0.100
Econometrica	69	6.931	(0.993)		0.037	0.296	0.878	0.525	0.089	0.000
Review of Economic Studies	63	6.922	(0.962)		0.034	0.353	0.863	0.478	0.107	0.036
Quarterly Journal of Economics	29	6.299	(1.306)		0.054	0.468	0.906	0.610	0.235	0.119
Journal of Monetary Economics	42	5.636	(1.150)		0.054	0.611	0.949	0.776	0.418	0.206
Journal of the European Economic Association	71	5.529	(0.970)		0.048	0.788	0.962	0.850	0.636	0.496
Journal of Political Economy	22	5.696	(1.728)		0.081	0.706	0.951	0.798	0.479	0.323
Journal of Econometrics	93	4.929	(0.525)		0.026	0.675	0.965	0.848	0.524	0.301
Journal of Development Economics	48	4.837	(1.314)		0.068	0.830	0.984	0.927	0.734	0.546
Journal of International Economics	36	4.615	(1.183)		0.060	0.816	0.973	0.890	0.688	0.487
Economic Journal	103	4.278	(0.551)		0.028	0.717	0.974	0.877	0.577	0.359
Journal of Public Economics	57	4.060	(0.756)		0.038	0.739	0.976	0.886	0.619	0.363
Journal of Applied Econometrics	24	4.033	(0.965)		0.047	0.681	0.959	0.829	0.527	0.246
Review of Economics and Statistics	59	3.897	(0.618)		0.030	0.665	0.968	0.849	0.508	0.227
Journal of Economic Theory	82	3.631	(0.541)		0.026	0.641	0.967	0.837	0.479	0.236
International Economic Review	28	3.194	(0.801)		0.038	0.617	0.957	0.801	0.443	0.222
Journal of Money, Credit and Banking	32	3.020	(0.700)		0.032	0.540	0.950	0.770	0.356	0.130
Games and Economic Behavior	78	2.980	(0.485)		0.022	0.488	0.955	0.782	0.312	0.050
European Economic Review	51	2.689	(0.559)		0.025	0.444	0.950	0.757	0.279	0.053
Economic Theory	48	2.641	(0.537)		0.024	0.536	0.944	0.726	0.312	0.127
Canadian Journal of Economics	24	2.466	(0.681)		0.030	0.450	0.936	0.711	0.254	0.046
Oxford Bulletin of Economics and Statistics	28	2.406	(0.733)		0.034	0.534	0.953	0.783	0.353	0.108
Econometric Theory	35	2.211	(0.741)		0.036	0.671	0.966	0.846	0.514	0.218
Journal of Health Economics	33	2.339	(0.636)		0.029	0.493	0.946	0.754	0.315	0.074
Journal of Economic Dynamics and Control	44	2.324	(0.671)		0.033	0.670	0.965	0.834	0.499	0.277
Oxford Economic Papers	24	1.959	(0.908)		0.044	0.634	0.967	0.840	0.454	0.288
Journal of Economic Behavior and Organization	42	1.755	(0.473)		0.020	0.429	0.932	0.695	0.221	0.021
Journal of Banking and Finance	23	1.050	(0.401)		0.014	0.246	0.859	0.492	0.063	0.014
Economics Letters	62	0.925	(0.277)		0.011	0.390	0.890	0.609	0.178	0.025
Other journals with less than 20 submissions										
Other journals with SJR ranking Top 100		4.206	(0.639)		0.031	0.584	0.958	0.813	0.425	0.171
Other journals with SJR ranking 101-1000		2.111	(0.211)		0.009	0.505	0.930	0.706	0.288	0.099
Other journals with SJR ranking 1001-3000		1.817	(0.184)		0.008	0.447	0.936	0.710	0.255	0.056
Other journals with SJR ranking 3001-6500		1.295	(0.170)		0.007	0.512	0.912	0.668	0.288	0.093
Other journals with SJR ranking Below 6500		0.691	(0.191)		0.008	0.428	0.881	0.609	0.187	0.027



cont..

Book series, unpublished work and other articles									
Economics & Econometrics: unpublished	4.252	(0.372)		0.018	0.640	0.969	0.839	0.470	0.210
Economics & Econometrics: book or book charpters	3.731	(0.841)		0.040	0.599	0.962	0.834	0.453	0.170
Economics & Econometrics: other articles	0.094	(0.547)		0.022	0.477	0.889	0.641	0.238	0.099
Cutoff Points									
Cutoff point 2	1.591	(0.072)		0.003	0.350	0.831	0.488	0.135	0.029
Cutoff point 3	4.476	(0.273)		0.013	0.691	0.971	0.850	0.528	0.286

SOURCE.—REF 2014 Economics and Econometrics Assessment Unit Results.

NOTE.—The results are obtained by 20,000 iterations after 3000 burn-in periods of the proposed MCMC algorithm. Column (1) lists the number of submission of each journal. Columns (2)-(4) summarise the posterior of the MCMC chain. Columns (5)-(10) present various MCMC diagnostic statistics. Model parameters are the parameters of the ordinal regression model (i.e. β and c)

Table 1.5: Estimated Rankings and Weighted Average Scores (Journals with More than 20 Submissions to REF Economics and Econometrics Panel)

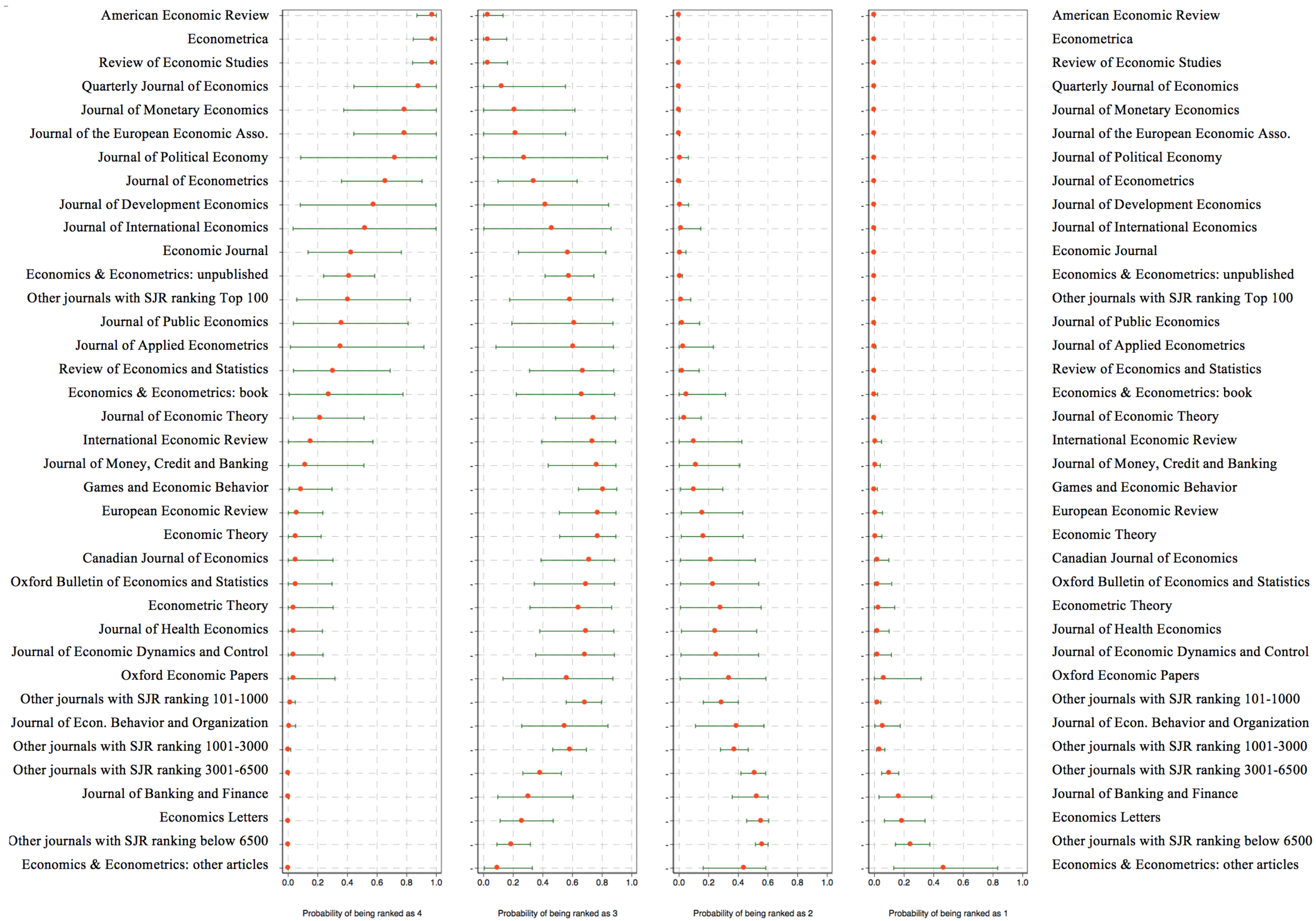
	No. of submiss-sions	SJR ranking 2014	Probability of being ranked as: 4		3		2		1*		Weighted average score	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Journals with more than 20 submissions												
American Economic Review	108	83	0.972	(0.037)	0.028	(0.037)	0.000	(0.000)	0.000	(0.000)	3.972	(0.037)
Econometrica	69	18	0.971	(0.043)	0.029	(0.043)	0.000	(0.000)	0.000	(0.000)	3.971	(0.043)
Review of Economic Studies	63	47	0.971	(0.047)	0.029	(0.047)	0.000	(0.000)	0.000	(0.000)	3.971	(0.047)
Quarterly Journal of Economics	29	6	0.877	(0.153)	0.122	(0.152)	0.000	(0.002)	0.000	(0.000)	3.877	(0.154)
Journal of Monetary Economics	42	292	0.786	(0.179)	0.213	(0.177)	0.001	(0.003)	0.000	(0.000)	3.786	(0.180)
Journal of the European Economic Association	71	134	0.784	(0.158)	0.215	(0.157)	0.001	(0.003)	0.000	(0.000)	3.784	(0.160)
Journal of Political Economy	22	20	0.722	(0.275)	0.272	(0.264)	0.006	(0.021)	0.000	(0.001)	3.716	(0.288)
Journal of Econometrics	93	290	0.658	(0.137)	0.340	(0.135)	0.002	(0.004)	0.000	(0.000)	3.657	(0.139)
Journal of Development Economics	48	439	0.577	(0.243)	0.416	(0.232)	0.007	(0.022)	0.000	(0.001)	3.569	(0.257)
Journal of International Economics	36	236	0.518	(0.286)	0.464	(0.264)	0.017	(0.044)	0.001	(0.003)	3.500	(0.316)
Economic Journal	103	358	0.363	(0.214)	0.614	(0.192)	0.023	(0.044)	0.001	(0.003)	3.339	(0.244)
Journal of Public Economics	57	463	0.357	(0.246)	0.609	(0.214)	0.033	(0.060)	0.001	(0.004)	3.322	(0.289)
Journal of Applied Econometrics	24	175	0.306	(0.173)	0.670	(0.152)	0.024	(0.038)	0.000	(0.002)	3.281	(0.201)
Review of Economics and Statistics	59	335	0.221	(0.130)	0.743	(0.107)	0.035	(0.042)	0.001	(0.002)	3.185	(0.164)
Journal of Economic Theory	82	375	0.156	(0.160)	0.740	(0.132)	0.099	(0.113)	0.006	(0.015)	3.045	(0.266)
International Economic Review	28	1218	0.115	(0.132)	0.766	(0.114)	0.113	(0.107)	0.006	(0.013)	2.991	(0.231)
Journal of Money, Credit and Banking	32	580	0.089	(0.078)	0.806	(0.064)	0.101	(0.075)	0.004	(0.006)	2.981	(0.149)
Games and Economic Behavior	78	1010	0.059	(0.062)	0.773	(0.095)	0.158	(0.112)	0.009	(0.015)	2.882	(0.186)
European Economic Review	51	883	0.053	(0.059)	0.769	(0.096)	0.168	(0.112)	0.010	(0.014)	2.866	(0.181)
Economic Theory	48	2818	0.052	(0.082)	0.712	(0.133)	0.218	(0.143)	0.019	(0.027)	2.797	(0.248)
Canadian Journal of Economics	24	3210	0.050	(0.082)	0.692	(0.151)	0.235	(0.157)	0.023	(0.033)	2.769	(0.272)
Oxford Bulletin of Economics and Statistics	28	285	0.041	(0.079)	0.645	(0.159)	0.282	(0.162)	0.033	(0.038)	2.693	(0.285)
Econometric Theory	35	917	0.038	(0.062)	0.696	(0.140)	0.244	(0.145)	0.022	(0.028)	2.751	(0.238)
Journal of Health Economics	33	2232	0.038	(0.061)	0.688	(0.145)	0.250	(0.149)	0.024	(0.031)	2.740	(0.247)
Journal of Economic Dynamics and Control	44	5894	0.036	(0.082)	0.560	(0.218)	0.336	(0.185)	0.068	(0.087)	2.565	(0.383)
Oxford Economic Papers	24	335	0.221	(0.130)	0.743	(0.107)	0.035	(0.042)	0.001	(0.002)	3.185	(0.164)
Journal of Economic Behavior and Organization	42	2934	0.008	(0.014)	0.549	(0.155)	0.388	(0.123)	0.055	(0.047)	2.510	(0.216)
Journal of Banking and Finance	23	2476	0.001	(0.002)	0.307	(0.132)	0.527	(0.062)	0.165	(0.093)	2.144	(0.223)
Economics Letters	62	5892	0.000	(0.001)	0.261	(0.093)	0.552	(0.037)	0.186	(0.071)	2.076	(0.162)
Other journals with less than 20 submissions												
Other journals with SJR ranking Top 100			0.406	(0.203)	0.581	(0.189)	0.013	(0.024)	0.000	(0.001)	3.392	(0.219)
Other journals with SJR ranking 101-1000			0.013	(0.012)	0.682	(0.060)	0.286	(0.059)	0.019	(0.010)	2.688	(0.085)
Other journals with SJR ranking 1001-3000			0.005	(0.005)	0.583	(0.060)	0.375	(0.049)	0.037	(0.015)	2.557	(0.078)
Other journals with SJR ranking 3001-6500			0.001	(0.001)	0.384	(0.066)	0.514	(0.043)	0.101	(0.029)	2.286	(0.095)
Other journals with SJR ranking Below 6500			0.000	(0.000)	0.190	(0.058)	0.562	(0.022)	0.249	(0.060)	1.941	(0.116)
Book series, unpublished work and other articles												
Economics & Econometrics: unpublished			0.415	(0.086)	0.580	(0.082)	0.006	(0.006)	0.000	(0.000)	3.409	(0.091)
Economics & Econometrics: book or book charpters			0.279	(0.221)	0.665	(0.182)	0.054	(0.087)	0.003	(0.013)	3.219	(0.294)
Economics & Econometrics: other articles			0.000	(0.001)	0.095	(0.090)	0.439	(0.118)	0.466	(0.189)	1.630	(0.272)

Source.—REF 2014 Economics and Econometrics Assessment Unit Results.

Note.—The results are obtained by 20,000 iterations after 3000 burn-in periods of the proposed MCMC algorithm. Columns (1)-(2) list the number of submission of each journal and the SJR 2014 journal rankings. Columns (3)-(10) present the posterior marginal effects of interest. Column (11)-(12) report the weighted average scores of journals and other types of publication. MCMC posterior standard deviations are reported in parentheses.

* The unclassified category is merged into the first category.

Figure 1.6: Error-Bar Plots of the Estimated Probabilities of Being Ranked as 4, 3, 2, 1 (Journals with More than 20 Submissions to REF Economics and Econometrics Panel)



Note.—From left to right each plot represents the estimated probabilities of being perceived as of quality 4, 3, 2 or 1, respectively. The *red dot* indicates the posterior means of the probabilities (partial effects), and the *green bar* indicates the 95% highest posterior densities of the probabilities (partial effects).

Table 1.6: Estimated Model Parameters, Rankings and Weighted Average Scores (Journals with More than 10 Submissions to REF Economics and Econometrics Panel)

	No. of submiss-sions	SJR ranking 2014	Model parameter		Probability of being ranked as:								Weighted average score	
					4		3		2		1*			
			Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Journals with 10-20 submissions														
Quantitative Economics	11	262	7.747	(1.936)	0.909	(0.141)	0.090	(0.140)	0.000	(0.002)	0.000	(0.000)	3.909	(0.141)
Journal of Economic History	12	2177	7.610	(2.095)	0.881	(0.189)	0.119	(0.188)	0.000	(0.003)	0.000	(0.000)	3.880	(0.190)
American Econ. Journal: Econ. Policy	12	114	7.475	(2.042)	0.876	(0.198)	0.121	(0.187)	0.003	(0.027)	0.000	(0.001)	3.873	(0.213)
American Econ. Journal: Micro	14	293	5.840	(2.185)	0.590	(0.380)	0.372	(0.334)	0.036	(0.098)	0.001	(0.006)	3.551	(0.448)
American Econ. Journal: Applied Econ	16	65	5.537	(1.885)	0.511	(0.334)	0.473	(0.315)	0.015	(0.039)	0.000	(0.001)	3.496	(0.357)
Journal of Law, Econ. & Organization	11	2857	5.217	(2.490)	0.446	(0.388)	0.480	(0.333)	0.070	(0.142)	0.004	(0.016)	3.368	(0.496)
Explorations in Economic History	11	3918	4.017	(2.181)	0.304	(0.346)	0.501	(0.322)	0.152	(0.227)	0.043	(0.112)	3.065	(0.661)
Journal of Labor Economics	13	246	3.685	(2.248)	0.238	(0.348)	0.512	(0.294)	0.225	(0.248)	0.025	(0.053)	2.963	(0.600)
Journal of the Royal Statistical Soc:A	19	2077	3.528	(1.722)	0.164	(0.272)	0.641	(0.252)	0.180	(0.197)	0.015	(0.038)	2.955	(0.454)
Scandinavian Journal of Economics	14	2665	3.562	(1.103)	0.140	(0.196)	0.725	(0.200)	0.128	(0.172)	0.007	(0.022)	2.997	(0.345)
Labour Economics	19	2674	3.275	(1.136)	0.103	(0.162)	0.714	(0.194)	0.171	(0.189)	0.012	(0.033)	2.908	(0.348)
Review of Economic Dynamics	14	295	2.902	(1.506)	0.080	(0.197)	0.635	(0.254)	0.258	(0.219)	0.027	(0.058)	2.767	(0.427)
Intl Journal of Game Theory	10	6681	2.910	(1.222)	0.066	(0.117)	0.685	(0.245)	0.210	(0.203)	0.039	(0.113)	2.777	(0.430)
Public Choice	14	2414	2.625	(1.156)	0.056	(0.124)	0.612	(0.231)	0.299	(0.225)	0.033	(0.060)	2.691	(0.395)
Journal of Mathematical Economics	18	3624	2.903	(0.824)	0.039	(0.072)	0.743	(0.170)	0.206	(0.172)	0.012	(0.029)	2.809	(0.259)
Intl Journal of Industrial Organization	16	1644	2.056	(1.250)	0.037	(0.123)	0.465	(0.279)	0.415	(0.239)	0.082	(0.097)	2.458	(0.464)
Theory and Decision	13	4266	2.515	(1.023)	0.037	(0.087)	0.619	(0.227)	0.311	(0.218)	0.032	(0.053)	2.661	(0.354)
RAND Journal of Economics	16	362	2.307	(1.294)	0.037	(0.087)	0.546	(0.287)	0.343	(0.238)	0.074	(0.118)	2.547	(0.470)
Environmental and Resource Econ.	15	1973	2.801	(0.864)	0.036	(0.069)	0.715	(0.187)	0.233	(0.188)	0.015	(0.031)	2.772	(0.278)
American Econ. Journal: Macro	11	159	0.971	(1.296)	0.033	(0.147)	0.182	(0.208)	0.515	(0.176)	0.270	(0.187)	1.978	(0.525)
Social Choice and Welfare	14	4073	2.688	(0.853)	0.030	(0.062)	0.693	(0.202)	0.258	(0.196)	0.019	(0.038)	2.735	(0.291)
Economic Inquiry	13	2161	2.542	(0.918)	0.027	(0.058)	0.650	(0.217)	0.297	(0.207)	0.026	(0.044)	2.677	(0.312)
Journal of Business & Econ. Statistics	13	242	1.902	(1.050)	0.024	(0.081)	0.458	(0.261)	0.432	(0.212)	0.087	(0.113)	2.418	(0.429)
Macroeconomic Dynamics	10	4771	2.052	(1.007)	0.020	(0.066)	0.503	(0.251)	0.414	(0.216)	0.063	(0.084)	2.479	(0.380)
Economica	18	2280	2.013	(1.138)	0.014	(0.042)	0.504	(0.292)	0.394	(0.222)	0.087	(0.124)	2.445	(0.436)
Manchester School	10	8203	1.271	(0.934)	0.003	(0.015)	0.312	(0.246)	0.511	(0.163)	0.174	(0.180)	2.145	(0.414)
J of Environ. Econ. & Management	11	744	1.108	(0.930)	0.003	(0.015)	0.274	(0.241)	0.515	(0.156)	0.209	(0.199)	2.071	(0.426)
Journal of Intl. Money and Finance	15	3066	1.297	(0.682)	0.001	(0.007)	0.306	(0.210)	0.551	(0.134)	0.141	(0.125)	2.168	(0.326)

cont..

Theoretical Economics	13	204	0.599	(0.575)	0.000	(0.001)	0.129	(0.121)	0.568	(0.102)	0.303	(0.169)	1.826	(0.276)
Journal of Agricultural Economics	10	3890	-0.928	(1.303)	0.000	(0.001)	0.028	(0.069)	0.262	(0.207)	0.710	(0.249)	1.318	(0.302)
Journal of Public Economic Theory	12	4020	-1.294	(1.213)	0.000	(0.000)	0.011	(0.028)	0.185	(0.179)	0.804	(0.200)	1.207	(0.223)
Journals with more than 20 submissions														
American Economic Review	108	83	7.616	(0.972)	0.968	(0.048)	0.032	(0.048)	0.000	(0.000)	0.000	(0.000)	3.968	(0.048)
Econometrica	69	18	7.516	(0.947)	0.966	(0.046)	0.034	(0.046)	0.000	(0.000)	0.000	(0.000)	3.966	(0.046)
Review of Economic Studies	63	47	7.359	(0.767)	0.963	(0.045)	0.037	(0.045)	0.000	(0.000)	0.000	(0.000)	3.963	(0.045)
Quarterly Journal of Economics	29	6	6.476	(1.141)	0.817	(0.155)	0.183	(0.155)	0.000	(0.000)	0.000	(0.000)	3.816	(0.155)
Journal of Monetary Economics	42	292	6.244	(1.301)	0.729	(0.238)	0.271	(0.237)	0.001	(0.003)	0.000	(0.000)	3.728	(0.240)
Journal of Political Economy	22	20	6.383	(1.572)	0.725	(0.255)	0.274	(0.253)	0.001	(0.005)	0.000	(0.000)	3.724	(0.257)
Journal of Econometrics	93	290	5.923	(0.731)	0.717	(0.152)	0.283	(0.152)	0.000	(0.000)	0.000	(0.000)	3.717	(0.153)
Journal of the European Econ. Asso.	71	134	5.788	(0.956)	0.653	(0.215)	0.346	(0.214)	0.001	(0.002)	0.000	(0.000)	3.652	(0.215)
Journal of Development Economics	48	439	5.548	(1.180)	0.582	(0.285)	0.412	(0.275)	0.006	(0.020)	0.000	(0.000)	3.576	(0.296)
Journal of International Economics	36	236	5.440	(1.627)	0.533	(0.356)	0.449	(0.332)	0.018	(0.042)	0.000	(0.001)	3.514	(0.383)
Journal of Public Economics	57	358	4.856	(1.174)	0.381	(0.275)	0.602	(0.257)	0.017	(0.035)	0.000	(0.001)	3.363	(0.296)
Journal of Applied Econometrics	24	463	4.547	(1.180)	0.334	(0.280)	0.624	(0.245)	0.041	(0.080)	0.001	(0.003)	3.292	(0.335)
Economic Journal	103	308	4.416	(0.712)	0.250	(0.168)	0.730	(0.151)	0.020	(0.039)	0.000	(0.001)	3.230	(0.192)
Review of Economics and Statistics	59	175	4.227	(0.695)	0.203	(0.159)	0.771	(0.141)	0.025	(0.037)	0.000	(0.001)	3.178	(0.183)
Journal of Economic Theory	82	335	3.785	(0.583)	0.102	(0.086)	0.848	(0.070)	0.049	(0.056)	0.001	(0.002)	3.053	(0.128)
Journal of Money, Credit and Banking	32	1218	3.448	(0.896)	0.087	(0.111)	0.790	(0.126)	0.119	(0.137)	0.004	(0.010)	2.960	(0.233)
International Economic Review	28	375	3.238	(1.003)	0.082	(0.144)	0.749	(0.159)	0.162	(0.159)	0.007	(0.014)	2.907	(0.282)
European Economic Review	51	1010	3.435	(0.692)	0.065	(0.077)	0.836	(0.087)	0.097	(0.097)	0.002	(0.005)	2.964	(0.160)
Games and Economic Behavior	78	580	3.409	(0.662)	0.062	(0.072)	0.839	(0.076)	0.098	(0.090)	0.002	(0.003)	2.961	(0.150)
Oxford Economic Papers	24	5894	2.829	(1.061)	0.047	(0.090)	0.690	(0.220)	0.240	(0.211)	0.024	(0.048)	2.760	(0.341)
Oxford Bulletin of Econ. and Statistics	28	3210	2.875	(0.823)	0.038	(0.079)	0.739	(0.148)	0.215	(0.159)	0.009	(0.016)	2.804	(0.234)
Canadian Journal of Economics	24	2818	2.776	(0.844)	0.033	(0.060)	0.714	(0.198)	0.238	(0.197)	0.016	(0.029)	2.764	(0.281)
Econometric Theory	35	285	2.696	(0.747)	0.031	(0.072)	0.703	(0.158)	0.254	(0.169)	0.012	(0.018)	2.754	(0.244)
Journal of Health Economics	33	917	2.486	(0.713)	0.017	(0.044)	0.663	(0.175)	0.301	(0.172)	0.018	(0.027)	2.680	(0.242)
Economic Theory	48	883	2.728	(0.612)	0.017	(0.025)	0.742	(0.151)	0.231	(0.153)	0.009	(0.015)	2.768	(0.194)
Journal of Econ. Dynamics & Control	44	2232	2.576	(0.657)	0.015	(0.031)	0.699	(0.162)	0.271	(0.161)	0.014	(0.019)	2.717	(0.212)
Journal of Econ.Behavior & Org.	42	2934	2.022	(0.668)	0.006	(0.019)	0.528	(0.206)	0.423	(0.179)	0.043	(0.045)	2.498	(0.264)
Journal of Banking and Finance	23	2476	1.283	(0.438)	0.000	(0.001)	0.287	(0.141)	0.593	(0.082)	0.120	(0.082)	2.168	(0.217)
Economics Letters	62	5892	1.131	(0.317)	0.000	(0.000)	0.234	(0.100)	0.625	(0.050)	0.141	(0.069)	2.094	(0.165)
Other journals with less than 10 submissions														
with SJR ranking top 100			4.000	(0.967)	0.188	(0.195)	0.751	(0.166)	0.060	(0.090)	0.001	(0.004)	3.126	(0.260)
with SJR ranking 101-1000			2.929	(0.614)	0.025	(0.036)	0.790	(0.108)	0.179	(0.119)	0.005	(0.009)	2.835	(0.158)
with SJR ranking 3001-6500			1.531	(0.274)	0.000	(0.001)	0.362	(0.096)	0.568	(0.066)	0.070	(0.037)	2.292	(0.130)

cont..

with SJR ranking 1001-3000	1.466	(0.236)	0.000	(0.000)	0.337	(0.083)	0.586	(0.058)	0.077	(0.032)	2.261	(0.113)
with SJR ranking below 6500	0.855	(0.274)	0.000	(0.000)	0.160	(0.074)	0.636	(0.034)	0.205	(0.072)	1.955	(0.143)
Book series, unpublished work and other												
Unpublished	4.922	(0.478)	0.385	(0.093)	0.612	(0.091)	0.003	(0.003)	0.000	(0.000)	3.382	(0.095)
Book or book charpters	3.992	(0.878)	0.170	(0.172)	0.777	(0.145)	0.052	(0.073)	0.001	(0.002)	3.116	(0.223)
Other articles	0.459	(0.715)	0.000	(0.001)	0.124	(0.136)	0.521	(0.135)	0.355	(0.219)	1.769	(0.339)
Cutoff points												
Cutoff point 2	5.224	(0.359)	---	---	---	---	---	---	---	---	---	---
Cutoff point 3	1.897	(0.114)	---	---	---	---	---	---	---	---	---	---

Source.—REF 2014 Economics and Econometrics Assessment Unit Results.

Note.—The results are obtained by 5000 iterations after 1000 burn-in periods of the proposed MCMC algorithm. Columns (1)-(2) list the number of submission of each journal and the SJR 2014 journal rankings. Columns (3)-(4) summarise the posterior of the MCMC chain. Columns (5)-(12) present the posterior marginal effects of interest. Columns (13)-(14) report the weighted average scores of journals and other types of publication. MCMC posterior standard deviations are reported in the parentheses. Model parameters are the parameters of the ordinal regression model (i.e. β and c)

* The unclassified category is merged into the first category.

Table 1.7: Estimated Model Parameters, Rankings and Weighted Average Scores (Economics Journals with More than 10 Submissions to REF Business and Management Studies Panel)

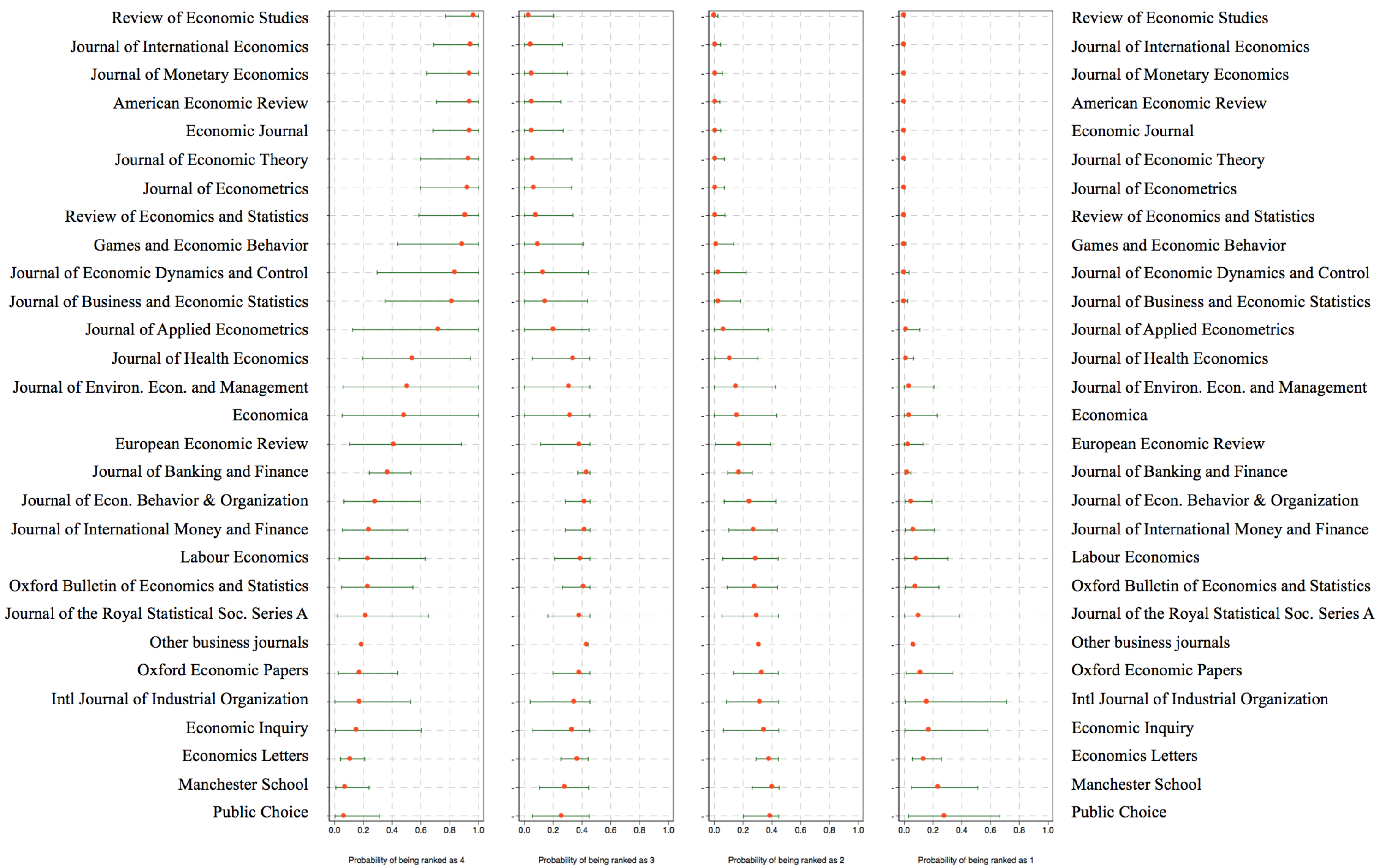
	No. of submiss-sions	SJR ranking 2014	Model parameter		Probability of being ranked as: 4		3		2		1*		Weighted average score	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Review of Economic Studies	13	47	5.798	(2.092)	0.967	(0.065)	0.030	(0.057)	0.003	(0.008)	0.000	(0.000)	3.964	(0.073)
Journal of International Economics	13	236	5.224	(1.658)	0.949	(0.088)	0.046	(0.075)	0.005	(0.013)	0.000	(0.001)	3.943	(0.102)
Journal of Monetary Economics	13	292	5.532	(2.295)	0.942	(0.102)	0.051	(0.084)	0.007	(0.019)	0.000	(0.002)	3.934	(0.122)
American Economic Review	28	83	4.834	(1.496)	0.939	(0.086)	0.055	(0.074)	0.006	(0.012)	0.000	(0.001)	3.933	(0.098)
Economic Journal	43	308	6.321	(4.096)	0.939	(0.094)	0.055	(0.081)	0.006	(0.014)	0.000	(0.001)	3.932	(0.109)
Journal of Economic Theory	11	335	5.122	(1.873)	0.930	(0.115)	0.061	(0.094)	0.008	(0.022)	0.000	(0.002)	3.921	(0.138)
Journal of Econometrics	26	290	5.129	(2.247)	0.923	(0.112)	0.068	(0.093)	0.009	(0.019)	0.000	(0.002)	3.913	(0.133)
Review of Economics and Statistics	23	175	4.596	(1.607)	0.907	(0.120)	0.082	(0.099)	0.011	(0.022)	0.001	(0.002)	3.895	(0.143)
Games and Economic Behavior	13	580	4.618	(1.768)	0.884	(0.163)	0.096	(0.122)	0.018	(0.041)	0.002	(0.011)	3.863	(0.214)
Journal of Econ. Dynamics & Control	18	2232	4.160	(1.616)	0.836	(0.194)	0.132	(0.139)	0.029	(0.055)	0.003	(0.010)	3.801	(0.260)
Journal of Business & Econ. Statistics	15	242	4.077	(1.688)	0.819	(0.195)	0.147	(0.142)	0.032	(0.052)	0.003	(0.008)	3.781	(0.257)
Journal of Applied Econometrics	14	463	3.573	(1.676)	0.723	(0.254)	0.202	(0.156)	0.064	(0.094)	0.011	(0.034)	3.637	(0.394)
Journal of Health Economics	24	917	2.520	(0.610)	0.540	(0.192)	0.338	(0.108)	0.109	(0.079)	0.013	(0.018)	3.404	(0.298)
J of Environ. Econ. & Management	12	744	2.631	(1.567)	0.503	(0.295)	0.308	(0.151)	0.153	(0.133)	0.036	(0.061)	3.279	(0.518)
Economica	16	2280	2.560	(1.430)	0.488	(0.288)	0.318	(0.148)	0.158	(0.131)	0.037	(0.064)	3.256	(0.512)
European Economic Review	24	1010	2.164	(0.651)	0.416	(0.197)	0.384	(0.089)	0.171	(0.102)	0.030	(0.035)	3.186	(0.356)
Journal of Banking and Finance	183	2476	2.048	(0.199)	0.372	(0.073)	0.431	(0.023)	0.175	(0.043)	0.022	(0.011)	3.152	(0.137)
J of Econ. Behavior & Organization	46	2934	1.750	(0.419)	0.280	(0.132)	0.421	(0.044)	0.246	(0.091)	0.053	(0.047)	2.927	(0.305)
Journal of Intl Money and Finance	40	3066	1.607	(0.403)	0.236	(0.117)	0.418	(0.045)	0.278	(0.088)	0.068	(0.054)	2.822	(0.302)
Labour Economics	19	2674	1.545	(0.553)	0.231	(0.157)	0.393	(0.068)	0.289	(0.109)	0.087	(0.079)	2.767	(0.403)
Oxford Bulletin of Econ. and Statistics	21	3210	1.571	(0.458)	0.230	(0.130)	0.408	(0.054)	0.285	(0.097)	0.076	(0.065)	2.793	(0.342)
J of the Royal Statistical Soc. Series A	14	2077	1.478	(0.613)	0.220	(0.164)	0.380	(0.081)	0.297	(0.114)	0.103	(0.103)	2.716	(0.449)
Oxford Economic Papers	21	5894	1.333	(0.470)	0.171	(0.113)	0.384	(0.073)	0.331	(0.090)	0.114	(0.087)	2.613	(0.358)
Intl Journal of Industrial Organization	17	1644	1.211	(0.735)	0.171	(0.144)	0.348	(0.115)	0.319	(0.106)	0.161	(0.175)	2.529	(0.523)
Economic Inquiry	18	2161	1.136	(0.730)	0.151	(0.156)	0.331	(0.112)	0.343	(0.105)	0.175	(0.157)	2.459	(0.505)
Economics Letters	76	5892	1.134	(0.234)	0.112	(0.043)	0.370	(0.049)	0.383	(0.042)	0.135	(0.051)	2.459	(0.184)
Manchester School	26	8203	0.780	(0.438)	0.072	(0.062)	0.285	(0.096)	0.406	(0.050)	0.238	(0.125)	2.190	(0.330)
Public Choice	14	2414	0.667	(0.577)	0.069	(0.078)	0.259	(0.117)	0.390	(0.065)	0.282	(0.171)	2.115	(0.419)
Other journals and publications			1.491	(0.014)	0.187	(0.003)	0.435	(0.003)	0.310	(0.004)	0.068	(0.002)	2.741	(0.008)
Cutoff points														
Cutoff point 2			1.180	(0.013)	---	---	---	---	---	---	---	---	---	---
Cutoff point 3			2.381	(0.016)	---	---	---	---	---	---	---	---	---	---

Source.—REF 2014 Business and Management Studies Assessment Unit Results.

Note.—The same as the notes for Figure 1.6.

* The unclassified category is merged into the first category.

Figure 1.7: Error-bar Plots of the Estimated Probabilities of Being Ranked as 4, 3, 2, 1 (Economics Journals with More than 10 Submissions to REF Business and Management Studies Panel)



Note.—The same as the notes for Table 1.6.

Chapter 2

Estimating Average Treatment Effects in Evaluation Studies: Using the Dirichlet Process Matching [1]

2.1 Introduction

Program evaluation studies typically conduct comparisons between treated and controlled units in order to learn about the treatment effect on individuals who have participated in a certain program. A prominent example is given by LaLonde (1986), which estimates the earnings effects of a US job training program by comparing the program participants (treatment group) and the non-participants (control group). If both groups were randomly selected into the study, simply comparing their average wages would have yielded an unbiased average treatment effect (ATE). However, with observational (or non-experimental) data, participants and non-participants may differ significantly along a number of important dimensions (i.e. participants are not randomly assigned to the treatment), and thus

[1] I would like to thank Sascha O. Becker, Mingli Chen, Clément de Chaisemartin, and Michael Pitt for helpful suggestions and conversations throughout this project. I would like to thank Wiji Arulampalam, Gianna Boero, François Caron, Bo E. Honoré, Koen Jochmans, Arthur Lewbel, Konrad Menzel, Jeremy Smith, Roland Rathlot, Mark Steel, and Massimiliano Tani for helpful conversations as well as seminar participants at the CAGE seminar, Warwick Annual Econometric Workshop, and Southampton PhD seminar for helpful feedback. This work is financially supported by PhD scholarship from the Department of Economics at the University of Warwick. All remaining errors are my own.

the controlled units are not capable of acting as an ideal counterpart to reflect the whole picture of the treated units. This constitutes a central problem in the program evaluation literature—the untreated outcome of the treatment group cannot be observed in reality.

Matching, aimed at equalizing or balancing the distribution of covariates across treatment and control groups (Stuart, 2010), has widely been adopted to address the "missing data" problem. Dehejia and Wahba (1999, 2002) demonstrate that when background characteristics of the treated and controlled groups are very distinct, matching is a more effective method to reduce the bias in the estimation of the average treatment effect on the treated (ATT), compared to conventional regression adjustments or selection models. Moreover, matching is a flexible approach that can be used in combination with other strategies (e.g. regression adjusted methods). The focus of this paper is on matching methodology. Unlike existing studies, I suggest an alternative matching estimator, which naturally meets the balancing property by construction, and hence attains a few desirable properties addressed in the literature. Moreover, the proposed matching estimator yields an ATT estimator that has an honest confidence band by taking account the errors from the matching.

Two most popular matching methodologies among researchers are arguably Mahalanobis matching and propensity score matching. These matching methods have been proved by Rubin and Thomas (1992a,b, 1996) and Rubin and Stuart (2006) to be equal-percent-bias-reducing (EPBR) after imposing necessary conditions on matching covariates.[2] In the absence of the EPBR condition, the propensity score matching can still be useful as it possesses sound theoretical properties. Indeed, the propensity score, which embodies the joint distribution of the underlying covariates, can serve as a "balancing score", if the score has been correctly specified. The propensity score is described as "balancing" if underlying covariates are balanced between the treatment group and the controlled group.[3]

[2]The EPBR condition requires that the matching covariates are (mixtures of) symmetric elliptical distributed. However, this is a very demanding assumption and might be violated when, for instance, researchers try to incorporate discrete or categorical variables.

[3]As noted by Abadie and Imbens (2016), "misspecification of the propensity score typically leads to inconsistency of the treatment effect estimator, unless the misspecified propensity score constitutes a balancing score."

Unfortunately, in the real word, it is almost impossible for us to know the true specification underlying the propensity score. A practical solution suggested by Rosenbaum and Rubin (1984) is to manually and iteratively attempt different specifications until the score is well balanced. Implementing such procedures can, however, be time-consuming and complicated. As a consequence, people might feel reluctant to do so. For instance, Diamond and Sekhon (2012) review articles published in famous economics journals during the 2000s and find that only about 32 percent of the empirical applications of matching have actually considered covariate balance. It should be noted that although there exist a number of other solutions to overcome lack of knowledge of the true propensity score,[4] they are unanimously based on large sample properties.

The present paper develops a *balancing* matching approach, namely the Dirichlet process (DP) matching, which by its nature satisfies the balancing property. To put it simple, the DP matching first approximates a joint distribution of detailed confounding variables in the treatment group and then matches each controlled unit to the treated if their confounding variables are in the same covariate space. The DP matching process should therefore properly impose the stochastic balancing property between the treatment and control groups.

More specifically, the joint distribution of the covariates is approximated by the Dirichlet process mixture (DPM) of normal distributions. If the mixture components are sufficiently large, a mixture of normal distributions can well approximate any arbitrary multivariate distribution. See Ferguson (1983), Rasmussen and Ghahramani (2002) and Rossi (2013) for finite/infinite mixture with continuous variables, and Norets and Pelenis (2012) for finite mixture with both continuous and discrete marginals. For example, macroeconomists commonly use a mixture of seven normals to approach the log chi-squared error term in a stochastic volatility model (Kim et al., 1998; Stock and Watson, 2007). It is found that under circumstances in which distributions have a waving shape, a mixture of normals with tiny component variances can precisely approximate the shape of the distributions

[4]For instance, one can adopt the safeguard options such as the bias-corrected estimator by Abadie and Imbens (2012), or utilize the nonparametric sieve method to approximate the true propensity score.

(Rossi, 2013). In addition, one might worry about the choice of the number of components in a finite mixture model. In such a case, the DPM of normals represents a more flexible alternative that explicitly allows for a countably infinite number of mixture components. This can be interpreted as a limit of the finite mixture of normals (Neal, 2000). Moreover, as is often the case, we need to match on characteristics that are usually measured on discrete scales, such as ethnicity, marital status, and gender. The discrete confounding variables in my matching model are augmented using Gaussian latent variables, following a standard Bayesian procedure proposed by Albert and Chib (1993) and later adopted in a setting of the mixture of normals model by Norets and Pelenis (2012).

The algorithm put forward by this paper functions as follows. It starts with approximating the covariate distribution of the treatment group by a DPM of normals. It subsequently utilizes another DPM of normals (for both the treatment and control groups) that fixes the covariate space of the treatment group at the posterior derived from the first DPM model, and distinguishes whether each controlled unit 1) is matched to one of the same mixtures as the treated; or (2) belongs to the same mixture as other unmatched controlled units; or (3) forms a new mixture distribution for this controlled unit *per se*. According to this matching design, the presence of the untreated sample should not affect the covariate space of the treated sample. The reason is that the mixture parameters for the matched controlled units are locked at their posteriors estimated using the first DPM that relies only on the treated units. Another representation of the DP is the Chinese restaurant process (CRP). This paper extends this metaphor to the DP matching. Specifically, I describe the DP matching strategy as a story of the *"Chinese restaurant with invited guests"*.

The DP matching approach presented in this paper contributes to the existing literature in three important aspects. First, as explained earlier, this method has the nice feature that covariate balancing condition is automatically fulfilled. Second, the DP matching, as a compromise of full covariates matching, is able to produce an efficient ATT estimator. Hahn (1998) analytically shows that whereas full covariates matching delivers an efficient ATT estimator, propensity score matching can merely be viewed as a dimensional reduction device for the ATT estimator. A related concern regarding efficiency of the estimator arises in Angrist and Hahn (2004). They use a simple example to empirically illustrate that, when matched cells are small or information of the treatment group is insufficient, matching on full covariates will be less efficient than stratification matching via a propensity score. This is because the stratification on the propensity score takes advantage of information from the full treatment sample, while the full covariates matching relies only on information about each cell. Fortunately, this is less of a problem in the DP matching as it exploits characteristics of the full treatment sample to form matched cells and mixtures. The DP matching method proposed here effectively combines the advantages of both full covariates matching and propensity score matching in the sense that it is a full covariates matching itself but also extracts complete information of the treatment group. As a result, the DP matching has great potential to induce an efficient ATT estimator.

Third, a further benefit of this new method lies in its ability to produce a credible confidence band. In fact, the traditional propensity score matching scheme is a two-step approach, with a propensity score being estimated in the first step and an ATT estimator being computed accordingly in the second step. Yet, Abadie and Imbens (2002, 2016) point out that errors in the estimation of the propensity score could lead to incorrect confidence intervals of the ATT estimator. This has motivated the use of bootstrap. However, existing research (e.g. Abadie and Imbens, 2008) has not established the credibility of the bootstrap standard errors of the ATT estimator. In the DP matching, the ATT estimator is integrated into a single Markov Chain Monte Carlo (MCMC) algorithm. As a result, the posterior standard deviations of the ATT estimator can be deemed as the "Bayesian equivalent" standard errors and can thus induce an honest confidence interval of the ATT estimator. Think of a simple one-dimensional covariate example (Figure 2.1). The controlled unit in red has a higher probability of being matched (classified into the mixture in blue) and has a lower probability of being unmatched (classified into the mixture in orange). Each MCMC iteration will classify this controlled unit into the mixture of the matched (unmatched) units with a higher (lower) probability. Unlike maximum likelihood estimation that seeks for an optimal classification, the MCMC samples are able to restore the probability of being matched or unmatched by sample, after performing sufficiently large iterations. An

ATT estimator is also calculated at each MCMC iteration step and can therefore directly reflect the errors in matching.

Based on simulation evidence and empirical results from applying the well-known LaLonde (1986) dataset, I conclude that the DP matching has several kinds of strength compared with Nearest Neighbour Matching via Mahalanobis distance or propensity score matching. Above all, the ATT estimator generated by the DP matching is robust to different specifications with different sets of confounding variables. In addition, a comparison of results from both the simulation and real data applications show that the DP matching can deliver a less biased ATT estimator relative to other matching methods. Last but not least, the DP matching results in a narrower interval estimator, as supported by the empirical application of the Lalonde data.

The remainder of the paper unfolds as follows. In Section 2.2, I formulate a treatment effect model framework and develop a Bayesian prospective of the treatment effect model. Section 2.3 presents details of the proposed DP matching method. In particular, I briefly review the DPM model and describe how the DPM of normals can approximate the covariate distribution of treatment group (Section 2.3.1), and then extend the DPM of normals to the DP matching by taking controlled units into consideration (Section 2.3.2). The data augmentation for the discrete part of the covariates is discussed in Section 2.3.3. Section 2.4 summarises the Gibbs sampling scheme for the DP matching. The empirical application[5] of the DP matching is presented in Section 2.5, where I rely on the benchmark LaLonde (1986) dataset in observational studies. Section 2.6 concludes this paper and discusses the potential limitations.

[5]Evidence from artificial data on comparisons between the DP matching and various propensity matching approaches is provided in Appendix B.1

2.2 Observational Studies and Average Treatment Effects

2.2.1 Notation

Let T_i denote the treatment indicator such that $T_i = 1$ when the i^{th} unit receives treatment, and $T_i = 0$ otherwise. Let X be the multidimensional vector of observable confounding variables (can be discrete and/or continuous variables). $Y_i(1)$ and $Y_i(0)$ respectively indicate the potential outcomes for the treated and the controlled units. For each individual i, we observe $Y_i = T_iY_i(1) + (1 - T_i)Y_i(0)$.

The observations $(i = 1, 2, ..., n)$ are randomly drawn from a relevant population. I use subscripts T and C to respectively indicate the treated and the controlled units. I organize observations so that the first n_T observations are associated with the treatment group (i.e. size of the treated units) and the remaining $n_C = n - n_T$ pertain to the control group (i.e. size of the controlled units). Similarly, $X_{i,T}$ and $X_{i,C}$ represent the confounding variables of the treated and untreated samples, respectively.

2.2.2 Framework and Estimands

The quantity of interest in the program evaluation is the average treatment effect on the treated (ATT). The *population* (PATT τ^p) and *sample* average treatment effect on the treated (SATT τ^s) are given by:

$$\tau^p := \mathrm{E}(Y_i(1) - Y_i(0)|T_i = 1); \qquad\qquad \tau^s := \frac{1}{n_T}\sum_{i:T_i=1}^{n}\Big(Y_i(1) - Y_i(0)\Big)$$

where $Y_i(1) - Y_i(0)$ is the unit-level treatment effect.[6] It is obvious that the unbiased estimator of the PATT (τ^p) is still the SATT (τ^s), and, therefore, a point estimator for

[6] As argued by Heckman and Robb (1985) and Heckman et al. (1997), the sub-population of the treated units is often of more interest than the overall population of the treated units in the context of narrowly targeted programs. In many evaluation studies, the sample is not randomly drawn from a well-defined population, resulting in difficulties in defining the population average treatment effect. In spite of this, one could still assume that this sample is randomly drawn from a relevant population, and then express the expected value of the SATT as an equivalent to the PATT (i.e. $\mathrm{E}(\tau^s) = \tau^p$).

the PATT should also be an appropriate point estimator for the SATT.

The PATT (τ^p) can be rewritten as $\mathrm{E}(Y_i(1)|T_i = 1) - \mathrm{E}(Y_i(0)|T_i = 1)$. Notice that conditional expectation $\mathrm{E}(Y_i(0)|T_i = 1)$ is counterfactual since we could not observe the untreated outcome of the treated units had they not been treated. In a randomized experimental design, individuals are randomly assigned to the treatment, i.e., $T \perp \big(Y(0), Y(1)\big)$. So one could simply replace $\mathrm{E}(Y_i(0)|T_i = 1)$ with the observable outcome of the controlled units $\mathrm{E}(Y_i(0)|T_i = 0)$. Conversely, when the treatment is not randomly assigned in observational studies, one solution is to assume that assignment to treatment is unconfounded (Rosenbaum and Rubin, 1983) and that the probability of receiving treatment is bounded away from zero to one. In algebra, $T \perp \big(Y(0), Y(1)\big)|X = x$ and $c < \Pr(T = 1|X = x) < 1 - c$ for some $c > 0$ and for all x in the support of $f_{X|T=1}$.[7] The estimand of the PATT can thus be specified as:

$$\begin{aligned}
\tau^p = \mathrm{E}(\tau^s) &= \int \mathrm{E}(Y(1) - Y(0)|X = x) f_{X|T=1}(x)\mathrm{d}x \\
&= \int \Big\{ \mathrm{E}(Y(1)|X = x, T = 1) \\
&\quad - \mathrm{E}(Y(0)|X = x, T = 0) \Big\} f_{X|T=1}(x)\mathrm{d}x \\
&= \int \tau_x f_{X|T=1}(x)\mathrm{d}x \qquad (2.2.1)
\end{aligned}$$

where the integral operator is over the support of X in a relevant population and $\tau_x := \Big\{ \mathrm{E}(Y(1)|X = x, T = 1) - \mathrm{E}(Y(0)|X = x, T = 0) \Big\}$ is the ATE for the sub-population with $X = x$.

2.2.3 Bayesian Inference with DP Matching

Next I build on the above facts in the literature, and develop an estimator of the ATT using a Dirichelet process matching method in a Bayesian fashion. I denote the total

[7] Therefore, instead of assuming random assignment of treatments, we assume that after conditioning on all observed covariates X, the treatment is as good as randomly assigned. Note that these two assumptions still hold when conditioning on the propensity score. See Rosenbaum and Rubin (1983) and Imbens (2000) for a rigorous proof.

number of matched mixture components by H which is indexed by $h = 1, ..., H$, and the resulting allocation of each unit by $\mathcal{S}_i \in \{1, 2, ..., H\}$. Conditioning on a suitable prior $\mathcal{H}$ and the data X, the DP matching procedure matches the treated and controlled units as well as generates a posterior of allocation $\mathcal{S}_i|X, \mathcal{H}$ for each unit. By construction, in each component h as defined by the DP matching, the covariates are balanced and the treatment assignment can be seen as random. Therefore, the ATE in a component τ_h, corresponding to τ_x in 2.2.1, can be calculated whilst considering the treatment as if it had been randomly assigned:

$$\tau_h := \mathrm{E}(Y(1)|T = 1, \mathcal{S} = h, X_h, \mathcal{H}) - \mathrm{E}(Y(0)|T = 0, \mathcal{S} = h, X_h, \mathcal{H}) \tag{2.2.2}$$

where X_h is the subsample of component h from the full sample of covariates X, and the expected values of $Y(1)$ and $Y(0)$ in the last equation are simply the posterior means of $Y(1)$ and $Y(0)$. Therefore, in the language of Bayesian modelling, the ATE can be regarded as the difference between the *posterior means* of the two groups; and the variance of the ATE is the variance of the difference between the *posterior means*. Assume that Y is normally distributed and I impose a conjugate normal-inverse-gamma prior on its parameters. By the conjugacy of normal likelihood and normal-inverse-gamma prior, the marginal posterior of τ_h follows a student-t distribution:[8]

$$\tau_h \sim \mathrm{t}_{v_n}\Big(\mathrm{E}(\tau_h|Y_h, \mathcal{S} = h, X_h, \mathcal{H}), \mathrm{Var}(\tau_h|Y_h, \mathcal{S} = h, X_h, \mathcal{H})\Big) \tag{2.2.3}$$

where v_n denotes the posterior degrees of freedom of the student-t distribution. It can be shown that, under a diffused prior, the marginal posterior mean of the ATE approaches the difference between the sample-averages of Y_T and Y_C:

$$\mathrm{E}(\tau_h|Y_h, \mathcal{S} = h, X_h, \mathcal{H}) \to \frac{\sum_{i:\mathcal{S}_i=h}^{n_T} Y_{i,T}}{n_{T,h}} - \frac{\sum_{i,C:\mathcal{S}_i=h}^{n_C} Y_{i,C}}{n_{C,h}} \tag{2.2.4}$$

[8] The textbook by Hoff (2006) provides the details of derivation, for example.

as the sample size goes to infinity. Here $n_{T,h}$ and $n_{C,h}$ are respectively the number of treated and controlled units in matched component h. Furthermore, the integral over the distribution of the covariates $f_{X|T=1}(x)$ in Equation (2.2.1) can be discretized into H parts by the DPM of normals, which implies that the posterior mean of the ATT conditional on the allocation can be written as:

$$\mathrm{E}(\tau|Y,\mathcal{S},X,\mathcal{H}) = \sum_{h=1}^{H}\frac{n_{T,h}}{n_T}\times \mathrm{E}(\tau_h|Y_h,\mathcal{S}=h,X_h,\mathcal{H}) \tag{2.2.5}$$

By the same token, the posterior variance of the ATT can be written as:

$$\begin{aligned}\mathrm{Var}(\tau|Y,\mathcal{S},X,\mathcal{H}) &= \frac{1}{n_T}\Big[\mathrm{Var}(Y|T=1,X,\mathcal{H}) \\ &\quad +\frac{1}{n_T}\sum_{h=1}^{H}\frac{n_{T,h}^2}{n_{C,h}}\mathrm{Var}(Y|T=0,\mathcal{S}=h,X_h,\mathcal{H})\Big]\end{aligned} \tag{2.2.6}$$

Note that the superscript τ has been eliminated. This is because the proposed estimator is not only used to estimate τ^p or $\mathrm{E}(\tau^s)$ but is also useful in the estimation of its sample analogue τ^s.

Finally, to feature the posterior distribution from the MCMC output, a Rao-Blackwell estimator of the ATT via the DP matching can be obtained as:

$$\widehat{\mathrm{E}}(\tau|Y,X,\mathcal{H}) = \frac{1}{J}\sum_{j=1}^{J}\mathrm{E}(\tau|Y,\mathcal{S}^{(j)},X,\mathcal{H}) \tag{2.2.7}$$

where $\mathcal{S}^{(j)}$ $(j=1,...,J)$ are the MCMC samples drawn from a target distribution $\mathcal{S}|X,\mathcal{H}$.

It is worth noting that, by the Rao-Blackwellization, the posterior variance of τ has the following representation:

$$\begin{aligned}\mathrm{Var}(\tau|Y,X,\mathcal{H}) &= \mathrm{E}_{\mathcal{S}|X,\mathcal{H}}\big[\mathrm{Var}(\tau|Y,\mathcal{S},X,\mathcal{H})\big] \\ &\quad +\mathrm{Var}_{\mathcal{S}|X,\mathcal{H}}\big[\mathrm{E}(\tau|Y,\mathcal{S},X,\mathcal{H})\big]\end{aligned} \tag{2.2.8}$$

Equation (2.2.8) reveals that the variance of the ATT consists of two parts: 1) the first term reflects the average variance of the ATT in the sample, and 2) the second term captures the variations trigged by the matching procedure. As highlighted in Section 2.1, errors in matching should be accounted for in the ATT (Abadie and Imbens, 2002, 2016), and the Bayesian MCMC method illustrated in this paper provides an automated solution by explicitly incorporating the matching variations into the variance estimator of the ATT.

2.3 Dirichlet Process Matching

2.3.1 Estimating the Distribution of Covariates of the Treated

This subsection describes how the DPM of normals model approximates the covariate distribution of the treatment group. This also serves as a brief review of the DP and DPM (see Teh, 2011; Teh et al., 2012, for a more rigorous review). For now, I let X be a random vector of continuous covariates. In Section 2.3.3 I extend X to include both continuous and discrete marginals.

The Dirichlet process was pioneered in Ferguson (1973). It refers to a distribution of the random probability measure P over a space Θ, such that for any finite measurable partition $A_1, ..., A_r$ of Θ, the vector $(P(A_1), ..., P(A_r))$ is distributed as a finite-dimensional *Dirichlet distribution* with parameter $(MG(A_1), ..., MG(A_r))$:

$$(P(A_1), ..., P(A_r)) \sim \mathsf{Dir}(MG(A_1), ..., MG(A_r)) \tag{2.3.1}$$

Then one could write $P \sim \mathsf{DP}(M, G)$ if P is a random probability measure with distribution given by the Dirichlet process. The Dirichlet process can be employed as a non-parametric prior on the parameters of a mixture model (Ferguson, 1983). Consider that we intend to model the distribution of covariates in the treatment group by the mixture of normals, then we could suppose that the data $X_{i,T}$ is generated through the mechanism below:

$$\begin{aligned} X_{i,T}|\theta_{i,T} &\sim F_{\mathcal{N}}(\theta_{i,T}) \\ \theta_{i,T}|P_T &\sim P_T \\ P_T &\sim \mathsf{DP}(M, G) \end{aligned} \tag{2.3.2}$$

Model (2.3.2) indicates that the covariates of the treatment group $X_{1,T}, ..., X_{n,T}$ are randomly drawn from a mixture of multivariate normal distributions $F_{\mathcal{N}}(\theta)$, where $\theta_T = (\mu_T, \Sigma_T)'$ includes parameters of the multivariate normal distribution. To formulate a mixture model, I let the mixing distribution or random probability measure on θ_T be P_T. The prior over P_T is distributed according a Dirichlet process, $\mathsf{DP}(M, G)$, with concentration parameter M and base distribution G. Hence, model (2.3.2) is formally referred as the *Dirichlet process mixture* (DPM) of normals.

The draws of θ_T from the DP are discrete and exhibit a clustering property. To see this, one needs to marginalize out P_T in model (2.3.2) and write the prior predictive distribution of $\theta_{i,T}$ in the following form (Blackwell and MacQueen, 1973):

$$\theta_{i,T}|\theta_{1,T}, ..., \theta_{i-1,T}, M, G \sim \frac{M}{M+i-1}G + \frac{\sum_{l=1}^{i-1} \delta_{\theta_{l,T}}}{M+i-1} \tag{2.3.3}$$

where δ_{θ_T} is the distribution concentrated at a single point θ_T. To understand the clustering property, one could consider the fact that $\theta_{1,T}, ..., \theta_{i-1,T}$ take only K_T distinct values say, $\theta^*_{1,T}, ..., \theta^*_{K_T,T}$. I thus rewrite (2.3.3) as:

$$\theta_{i,T}|\theta_{1,T}, ..., \theta_{i-1,T}, M, G \sim \frac{M}{M+i-1}G + \frac{\sum_{k=1}^{K_T} n_{k,T}}{M+i-1}\delta_{\theta^*_{k,T}} \tag{2.3.4}$$

where $n_{k,T}$ is the number of $\theta_{i,T}(i = 1, ..., n_T)$ that are equal to $\theta^*_{k,T}$. Obviously, the weighted average (i.e. with relevant proportions $\frac{M}{M+i-1}$ in the first term and $1 - \frac{M}{M+i-1}$ in the second term) in Equation (2.3.4) represents a mixture. Another perspective taken on the DP is the *Chinese restaurant process* (CRP) (Aldous, 1985; Pitman, 2002). It is helpful to be precise about the story behind the CRP. It describes a Chinese restaurant with an infinite number of tables: a customer $\theta_{i,T}$ enters the restaurant and will be stochastically seated according to the number of people who have already settled at each table. S/he can be seated at table $\theta^*_{k,T}$ with a probability proportional to the number of existing customers $n_{k,T}$ at this table, or be seated at a free table with a probability proportional to M.

The CRP representation not only features the clustering property but also helps to render an efficient collapsed Gibbs sampling scheme for the DPM of normals with a conjugate prior. Model (2.3.2) can be rewritten using the representation of the CRP mixture of normals:

$$\begin{aligned} X_{i,T}|\theta^*_{k,T} &\sim F_{\mathcal{N}}(\theta^*_{\mathcal{S}_{i,T}}) \\ \theta^*_{k,T}|\theta_0 &\sim G(\theta_0) \quad (k = 1, ...\infty) \\ \mathcal{S}_{i,T}|M, \mathcal{S}_{-i,T} &\sim \mathsf{CRP}(M, \mathcal{S}_{-i,T}) \quad (i = 1, ..., n_T) \end{aligned} \tag{2.3.5}$$

where $G(\theta_0)$ is chosen to be a natural conjugate prior distribution (Normal-Inverse-Wishart) with hyper-parameter θ_0. $\mathcal{S}$ denotes the seating arrangement or the mixture label for each observation, and is distributed as a Chinese restaurant process:[9]

$$p(\mathcal{S}_{i,T} = k|M, \mathcal{S}_{-i,T}) = \begin{cases} \dfrac{n_{k,T}}{M+i-1}, & k = 1, ..., K_T \\ \dfrac{M}{M+i-1}, & k > K_T \end{cases}$$

To sample $(\theta^*_{k,T}, \mathcal{S}_{i,T})$ from their posterior densities in model (2.3.5) when conjugate priors are used, the collapsed Gibbs sampling method can be adopted (Algorithm 2 in Neal (2000)). According to the Rao-Blackwell theorem, one could collapse out $\theta^*_{k,T}$ and sample

[9]See for example, Wood and Black (2008) for the derivation.

only $\mathcal{S}_{i,T}$ from its posterior distribution.[10] Therefore, in our case $S_{i,T}$ can be sampled by:

$$\begin{aligned}
& p(\mathcal{S}_{i,T}=k|\mathcal{S}_{-i,T},X_T,M,\theta_0) \\
\propto\ & p(\mathcal{S}_T|M)p(X_T|S_T,\theta_0) \\
=\ & p(\mathcal{S}_{i,T}=k|\mathcal{S}_{-i,T},M)p(X_{i,T}|\mathcal{S}_{i,T}=k,\mathcal{S}_{-i,T},X_{-i,T},M,\theta_0) \\
=\ & p(\mathcal{S}_{i,T}=k|\mathcal{S}_{-i,T},M)p(X_{i,T}|X_{k_{-i},T},\theta_0) \\
=\ & \frac{n_{k_{-i},T}}{M+n_T-1}\int F_{\mathcal{N}}(X_{i,T}|\theta^*_{k,T})p(\theta^*_{k,T}|X_{k_{-i},T},\theta_0)\mathrm{d}\theta^*_{k,T}
\end{aligned} \tag{2.3.6}$$

where $-i,T$ indicates all observations except for the current i, k_{-i} means the observations in the cluster k except for i, and the integral is essentially the posterior predictive density of $X_{i,T}$ based on prior $G(\theta_0)$ and all observations of X_T in component k except for the current i.

Since a conjugate prior of G (i.e., Normal-Inverse Wishart) is employed, the resulting posterior predictive turns out to follow a multivariate student-t distribution. Moreover, the update of S_i when i starts a new component or table is:

$$\begin{aligned}
& p(\mathcal{S}_{i,T}=K_T+1|\mathcal{S}_{-i,T},X_{i,T},\{\theta^*_{k,T}\}_1^\infty,M,\theta_0) \\
\propto\ & p(\mathcal{S}_{i,T}=K_T+1|\mathcal{S}_{-i,T},\{\theta^*_{k,T}\}_1^\infty,M,\theta_0)p(X_{i,T}|\mathcal{S}_{-i,T},M,\theta_0,\mathcal{S}_{i,T}=K_T+1) \\
=\ & p(\mathcal{S}_{i,T}=K_T+1|\mathcal{S}_{-i,T},M)p(X_{i,T}|\theta_0) \\
=\ & \frac{M}{M+n_T-1}\int F_{\mathcal{N}}(X_{i,T}|\theta^*_T)\mathrm{d}G(\theta^*_T|\theta_0)
\end{aligned} \tag{2.3.7}$$

where the latter term in the integral is merely the prior density of $X_{i,T}$.

2.3.2 Extension to the Dirichlet Process Matching

Next, I bring the control group into the picture and extend the above steps to a matching procedure. Remember that model (2.3.2) generates the posterior of $\{\theta_{i,T}, i=1,...,n_T\}$ of the treatment sample. Once the controlled units are included, I regard $\theta_{i,T}$ as known

[10] In simple words, the Rao-Blackwell theorem states that when sampling from $P(A,B|H)$, a more efficient way is to marginalize out B and sample only one conditional posterior $P(A|H)$. With a smaller state space, the sampler typically converges faster to the equilibrium distribution (Liu, 2001).

and fix it at the posterior, and then predict $\{\theta_{i,C}, i = n_T + 1, ..., n\}$ of the control sample. Therefore, the *Dirichlet process matching* could be formulated according to the following data generating process:

$$
\begin{aligned}
&\text{For } i = 1, ..., n_T : \\
&\qquad X_{i,T}|\theta_{i,T} \sim F_{\mathcal{N}}(\theta_{i,T}) \\
&\qquad \theta_{i,T}|P_T \sim P_T \\
&\qquad P_T \sim \mathsf{DP}(M, G)
\end{aligned}
\tag{2.3.8}
$$

$$
\begin{aligned}
&\text{For } i = 1, ..., n_T, n_T + 1, ..., n : \\
&\qquad X_i|\theta_i \sim F_{\mathcal{N}}(\theta_i) \\
&\qquad \theta = (\theta_T^{known}, \theta_C)' \\
&\qquad \theta_i|P \sim P \\
&\qquad P \sim \mathsf{DP}(M, G)
\end{aligned}
\tag{2.3.9}
$$

In model (2.3.9), $\{\theta_{i,T}, i = 1, ..., n_T\}$ of the treated units are treated as known and are fixed at their posteriors that have been obtained with model (2.3.8). For the controlled units, $\theta_{i,C}$ could either be equal to a certain $\theta_{i,T}$ that is the same as the treatment group, or be equal to other values that differ from the treatment group. If a controlled unit's covariates parameter $\theta_{i,C}$ is equal to any value of $\theta_{i,T}$, this unit will not contribute to the posterior θ_T, as θ_T is always fixed at the posterior using the treatment sample only. This is in line with the nature of matching: the successfully paired controlled and treated units are supposed to have the same joint distribution, while controlled units in their own rights exert trivial influences on the covariate distribution of treated units.

In model (2.3.9), the measurable space Θ can be partitioned into $r < \infty$ subsets $A_1, ..., A_{K_T}, A_{K_{T+1}}, ..., A_r$, with the following random probability measures:

$$P(A_1), ..., P(A_{K_T}), P(A_{K_{T+1}}), ..., P(A_r) \sim \mathsf{Dir}\big(MG(A_1), ..., MG(A_{K_T}), MG(A_{K_{T+1}}), ..., MG(A_r)\big)$$

Let $n_k = \#\{i : \theta_i \in A_k\}$ indicate the total count of observed values in partition A_k. Then the likelihood model is a multinomial distribution that contains the numbers of observed values $(n_1, ..., n_{K_T}, n_{K_T+1}, ..., n_r)$ for each subset. By the conjugacy of multinomial and Dirichlet distributions, we can derive the following Dirichlet posterior:

$$\begin{aligned}
&\big(P(A_1), ..., P(A_{K_T}), P(A_{K_{T+1}}), ..., P(A_r)\big)|\theta_1, ...\theta_{n_T}, \theta_{n_T+1}, ..., \theta_n \\
\sim\ & \mathsf{Dir}\Big(MG(A_1) + n_1, ..., MG(A_{K_T}) + n_{K_T}, MG(A_{K_{T+1}}) + n_{K_T+1}, ..., MG(A_r) + n_r\Big) \\
\sim\ & \mathsf{Dir}\Big(MG(A_1) + n_{1,C} + n_{1,T}, ..., MG(A_{K_T}) + n_{K_T,C} + n_{K_T,T}, \\
& MG(A_{K_{T+1}}) + n_{K_T+1,C} + n_{K_T+1,T}, ..., MG(A_r) + n_{r,T} + n_{r,C}\Big)
\end{aligned}$$

where the last line exploits the fact that n_k comprises numbers from both the treated and controlled units in partition k (i.e. $n_k = n_{k,T} + n_{k,C}$). By the definition of the Dirichlet process described earlier (Ferguson, 1973), this could be written as:

$$P|\theta_1, ...\theta_{n_T}, \theta_{n_T+1}, ..., \theta_n \quad \sim \quad \mathsf{DP}\Big(M + n_C + n_T, \frac{M}{M + n_C + n_T}G + \frac{\sum_{i=1}^{n_C+n_T} \delta_{\theta_i}}{M + n_C + n_T}\Big)$$

The predictive distribution of θ_{n+1} given observed $\theta_1, ..., \theta_n$, can be written as:[11]

$$\theta_{n+1}|\theta_1, ...\theta_{n_T}, \theta_{n_T+1}, ..., \theta_n \quad \sim \quad \frac{MG + \sum_{i=1}^{n_C+n_T} \delta_{\theta_i}}{M + n_C + n_T}$$

After grouping the same values of θ_i into every single θ_k^*, the CRP representation takes the following form:

$$\theta_{n+1}|\theta_1, ...\theta_{n_T}, \theta_{n_T+1}, ..., \theta_n \quad \sim \quad \frac{M}{M + n_C + n_T}G + \frac{\sum_{k=1}^{K_T+K_C} n_k \delta_{\theta_k^*}}{M + n_C + n_T}$$

[11] By the definition of the DP, the base distribution G is also the expected value of the process. Therefore, for any measurable partition A, the posterior mean of the base distribution is:

$$\mathrm{E}(P(A)|\theta_1, ..., \theta_n) \quad = \quad \frac{MG(A) + \sum_{i=1}^{n_C+n_T} \delta_{\theta_i}(A)}{M + n_C + n_T}$$

Since $\Pr(\theta_{n+1} \in A|\theta_1, ..., \theta_n) = \mathrm{E}(P(A)|\theta_1, ..., \theta_n)$, the predictive of $\theta_{n+1} \in A|\theta_1, ..., \theta_{n+1}$ can be viewed as a posterior of the base distribution.

Note that among all of the counts $(n_1, ..., n_{K_T}, n_{K_T+1}..., n_r)$, the first $(n_1, ..., n_{K_T})$ may simultaneously contain treated and controlled individuals, whereby the remaining $(n_{K_T+1}, ..., n_r)$ include only the control sample. This leads to the following representation:

$$
\begin{aligned}
& \theta_{n+1}|\theta_1, ...\theta_{n_T}, \theta_{n_T+1}, ..., \theta_n \\
\sim \; & \frac{M}{M + n_C + n_T} G + \frac{\sum_{k=1}^{K_T+K_C} n_k \delta_{\theta_k^*}}{M + n_C + n_T} \\
\sim \; & \frac{M}{M + n_C + n_T} G + \frac{\sum_{k=1}^{K_T+K_C} (n_{k,T} + n_{k,C}) \delta_{\theta_k^*}}{M + n_C + n_T} \\
\sim \; & \frac{M}{M + n_C + n_T} G + \frac{\sum_{k=1}^{K_T} (n_{k,T} + n_{k,C}) \delta_{\theta_{k,T}^{*known}}}{M + n_C + n_T} \\
& + \frac{\sum_{k=K_T+1}^{K_T+K_C} n_{k,C} \delta_{\theta_{k,C}^*}}{M + n_C + n_T}
\end{aligned}
\tag{2.3.10}
$$

Still, the expression in 2.3.10 is a mixture. It is important to note that for a particular controlled unit, there are three options available rather than two: 1) its $\theta_{i,C}$ is equal to the posterior of $\theta_{k,T}^{*known}$ obtained using the treatment sample only with a probability proportional to $n_{k,T}+n_{k,C}$ (*matched*); 2) its $\theta_{i,C}$ is equal to $\theta_{k,C}^*$ with a probability proportional to $n_{k,C}$ (*unmatched*); and 3) $\theta_{i,C}$ samples a new value from G with a probability proportional to M (*unmatched*).

I describe the DP matching as a story of the *Chinese restaurant with invited guests*: n_T hosts (the treated) came to a Chinese restaurant and sat at K_T tables. They have ordered $\theta_{k,T}^{*known}$ dishes, and are waiting for their invited guests (the controlled) to arrive and share the dishes. The controlled unit or the invited guest could be seated on one of the existing tables occupied by the host (the treated) and enjoy the dishes (posterior of $\theta_{k,T}^{*known}$). However, in the Chinese folk culture, it is the host who determines what to eat and the guest(s) rarely order food, suggesting that guests should not contribute to the values of $\theta_{k,T}^*$'s posterior. In cases where a controlled unit or a guest is not invited, s/he could sit at an existing table with other uninvited customers and order new dishes $\theta_{k,C}^*$; alternatively, s/he could sit at a new table. In this metaphor, only controlled units who are described as the *invited* guests, are considered as matched.[12]

Specifically, the CRP version of the DP matching has the following expression:

$$\begin{aligned}
&\text{For } i = 1, ..., n_T: \\
X_{i,T}|\theta^*_{k,T} &\sim F_{\mathcal{N}}(\theta^*_{\mathcal{S}_{i,T}}) \\
\theta^*_{k,T}|\theta_0 &\sim G(\theta_0) \quad (k = 1, ...\infty) \\
\mathcal{S}_{i,T}|M, \mathcal{S}_{-i,T} &\sim \mathsf{CRP}(M, \mathcal{S}_{-i,T}) \quad (i = 1, ..., n_T)
\end{aligned} \tag{2.3.11}$$

$$\begin{aligned}
&\text{For } i = 1, ..., n_T, n_T + 1, ..., n: \\
X_i|\theta_i &\sim F_{\mathcal{N}}(\theta_i) \\
\theta^* &= (\theta^{*known}_T, \theta^*_C)' \\
\theta^*_k|\theta_0 &\sim G(\theta_0) \quad (k = 1, ...\infty) \\
\mathcal{S} &= (\mathcal{S}^{known}_T, \mathcal{S}_C)' \\
\mathcal{S}_i|M, \mathcal{S}_{-i} &\sim \mathsf{CRP}(M, \mathcal{S}_{-i}) \quad (i = 1, ..., n_T)
\end{aligned} \tag{2.3.12}$$

In model (2.3.12), $\{\theta_{k,T}, k = 1, ..., K_T\}$ and $\{\mathcal{S}_{i,T}, i = 1, ..., n_T\}$ of the treated sample are held at their posteriors computed from model (2.3.11). Using the same argument of the collapsed Gibbs sampling as in the previous subsection, $\{\mathcal{S}_{i,C}, i = 1, ..., n_C\}$ for the control group can be updated from:

$$\mathcal{S}_{i_C}|\mathcal{S}_{-i}, X, M, \theta_0 \sim \begin{cases} \dfrac{n_{k,T} + n_{k_{-i},C}}{M + n_T + n_C - 1} \int F_{\mathcal{N}}(X_{i,C}|\theta^*_{k,T}) p(\theta^{*,known}_{k,T}|X_{k,T}, \theta_0) \mathrm{d}\theta^{*,known}_{k,T} \\ \dfrac{n_{k_{-i},C}}{M + n_T + n_C - 1} \int F_{\mathcal{N}}(X_{i,C}|\theta^*_{k,C}) p(\theta^*_{k,C}|X_{k_{-i},C}, \theta_0) \mathrm{d}\theta^*_{k,C} \\ \dfrac{M}{M + n_T + n_C - 1} \int F_{\mathcal{N}}(X_{i,C}|\theta^*_C) \mathrm{d}G(\theta^*_C|\theta_0) \end{cases}$$

where the integral in the first line corresponds to the predictive density of $X_{i,C}$ based on the posterior of $\theta^{*,known}_{k,T}|X_{k,T}, \theta_0$.

[12]Figures B.5-B.6 in the appendix to the paper visualize this metaphor

2.3.3 With Discrete Covariates

To deal with the discrete component of the covariates, I apply the classic Bayesian data augmentation approach by Albert and Chib (1993), which was originally developed to augment the dependent variable in a probit model. This approach basically augments discrete variables by the Gaussian latent variables. I denote the observed pre-treatment characteristics by $X \in \mathbb{R}^{q+p}$. The normal and the discrete part of X are denoted by $X^C \in \mathbb{R}^q$ and $X^D \in \mathbb{R}^p$, respectively. The Gaussian latent variable used to expand the discrete X_i^D is X^{D^*}. The distribution of X can be specified as a single normal distribution:

$$F_{\mathcal{N}}(X^C, X^{D^*}|\theta) = F_{\mathcal{N}}(X^C|\theta)F_{\mathcal{N}}(X^{D^*}|X^C, \theta)$$

where $\theta = (\theta^C, \theta^D)' = (\mu^C, \Sigma^C, \mu^D, \Sigma^D)'$. This suggests that when we have mixed-type data that has both Gaussian and discrete marginals, the discrete part can be treated as a Gaussian marginal and can be framed into a single multivariate Gaussian distribution. For instance, Pitt et al. (2006) and Smith and Khaled (2012) successfully augment discrete marginals and fit them into a single Gaussian copula. Norets and Pelenis (2012) employs this method to combine discrete and continuous variables into a single finite mixture of normal distributions.

With regard to posterior sampling, I am able to remain within the efficient collapsed Gibbs sampler with the aid of the Gaussian latent variable. In practice, one just needs to include an additional Gibbs step for updating the latent variable X^{D^*}. First, the conditional density for the latent variable $[X^{D^*}|\theta^*, X_i^C, X_i^{D^*}]$ in one of the mixture components[13] can be derived as follows:

$$\Pr(X_i^D = x|\theta^*, X_i^C, X_i^{D^*}) = \begin{cases} 1, & X_i^{D^*} \in (c^-_{X_i^D}, c^+_{X_i^D}] \\ 0, & \text{otherwise} \end{cases}$$

where $c^-_{X_i^D}$ and $c^+_{X_i^D}$ are lower and upper bounds of $X_i^{D^*}$ when transferring $X_i^D = x$ to

[13]Note that I omitted the mixture class label (subscription k on each variable) in following expressions for clarity.

$X_i^{D^*}$. Second, the individual value of the latent variable $X^{D^*} \in \mathbb{R}^p$ can be generated from:

$$
\begin{aligned}
f(X^{D^*}|\theta, X^D, X^C) &= \prod_{i=1}^{n} f(X_i^D|\theta^*, X_i^C, X_i^{D^*}) f_{\mathcal{N}}(X_i^{D^*}|\theta^*, X_i^C) \\
&= \prod_{i=1}^{n} \Pr(X_i^D = x|\theta^*, X_i^C, X_i^{D^*}) f_{\mathcal{N}}(X_i^{D^*}|\theta^*, X_i^C) \quad (2.3.13)
\end{aligned}
$$

where $f_{\mathcal{N}}(X_i^{D^*}|\theta^*, X_i^C) = \mathcal{N}(X_i^{D^*}|\mu^{DC}, \Sigma^{DC})$ in which, by the bivariate normal partition formula,

$$
\begin{aligned}
\mu^{DC} &= \mu^D + \Sigma^{DC} {\Sigma^{C}}^{-1} (X^C - \mu^C) \\
\Sigma^{DC} &= \Sigma^{DD} - \Sigma^{DC} {\Sigma^{C}}^{-1} \Sigma^{CD} \quad (2.3.14)
\end{aligned}
$$

It is worthwhile to mention again that the above conditional density should be generated repeatedly for each mixture component.

On top of this, by the nature of matching, the controlled sample should have no impact on the treatment group. Hence, X_T of the treatment sample is updated in the model (2.3.11) with treated units only. When it comes to the model (2.3.12), X_T is also locked at the posterior and the sampling of X would only update the X_C of the controlled units.

2.4 Estimation Method

2.4.1 Gibbs Sampler

Next, I combine above results and summarize the Gibbs sampling scheme for the proposed DP matching algorithm. Here I slightly change the definition of covariates X. Let X contains the continuous component and the latent Gaussian component $X = (X^C, X^{D^*})$. In order to assign a “default prior” on the covariates with different scales, I normalize each continuous marginal of X by subtracting its sample mean of treatment group and dividing by the treatment standard deviation. In each iteration $j = 1, ..., J$ of the Gibbs sampling, I sequentially update following conditional posteriors.

Step 1. Sampling the class labels of the treatment group $\mathcal{S}_{i,T}$

First, I focus on the treatment sample only (model 2.3.11). For the treatment sample $i = 1, ..., n_T$:

$$\mathcal{S}_{i,T}|\mathcal{S}_{-i,T}, X, M, \theta_0 \sim \begin{cases} \dfrac{n_{k_{-i},T}}{M + n_T - 1} \int F_{\mathcal{N}}(X_{i,T}|\theta^*_{k,CT}) p(\theta^*_{k,T}|X_{k_{-i},T}, \theta_0) \mathrm{d}\theta^*_{k,T} \\ \dfrac{M}{M + n_T - 1} \int F_{\mathcal{N}}(X_{i,T}|\theta^*_T) \mathrm{d}G(\theta^*_T|\theta_0) \end{cases}$$

If as a result of sampling $\mathcal{S}_{i,T}$ some mixture component k of the treated sample becomes empty (i.e. $n_{k,T} = 0$) in j^{th} iteration, then the previously matched controlled units to this k will be regarded as unmatched. And these controlled units themselves will form an additional mixture of unmatched units only.

Step 2. Sampling the mixture parameter of the treatment group

Using the collapsed Gibbs sampling, there is no need to update θ^*_k. However, to proceed to sampling from other conditionals, I still update θ^*_k. Then for each component $k = 1, ..., K_T$:

$$\theta^*_{k,T}|X_T, \theta_0 \sim \prod_{i:i \in k, T_i = 1} F_{\mathcal{N}}(X_{i,T}|\theta^*_{k,T}) G(\theta^*_{k,T}|\theta_0)$$

Step 3. Sampling the Gaussian latent variable of the treatment group

In the treatment sample, for each *discrete* marginal $X^{(l)}$ out of *all* $X = (X^{(l)}, X^{(-l)})$ and for each component $k = 1, ..., K_T$, compute $\Sigma^{(l)(-l)}_{k,T}$ and $\mu^{(l)(-l)}_{k,T}$ using formulas (2.3.14), and sample $X^{(l)}_T$ from following truncated normal:

$$X^{(l)}_T|\theta^*_{k,T}, X^{(l)D}_{k,T}, X^{(-l)}_{k,T} \sim \prod_{i:i \in k, T_i = 1} \Pr(X^{(l)D}_{i,T} = x|\theta^*_{k,T}, X^{(-l)}_{i,T}, X^{(-l)}_{i,T}) \mathcal{N}(X^{(l)}_{i,T}|\mu^{(l)(-l)}_{k,T}, \Sigma^{(l)(-l)}_{k,T})$$

Step 4. Sampling the class labels of the control group $\mathcal{S}_{i,C}$

Then I consider the model for matching (model 2.3.12), in which both treated units and controlled units are included. Note that $\{S_{i,T}, i = 1, ..., n_T\}$ are fixed at their postreriors from model (2.3.11). So I just need to update the class labels of the control group. For the control sample $i = n_T + 1, ..., n$:

$$\mathcal{S}_{i,C}|\mathcal{S}_{-i}, X, M, \theta_0 \sim \begin{cases} \dfrac{n_{k,T} + n_{k_{-i},C}}{M + n_T + n_C - 1} \int F_{\mathcal{N}}(X_{i,C}|\theta^*_{k,T})p(\theta^*_{k,T}|X_{k,T}, \theta_0)\mathrm{d}\theta^*_{k,T} \\ \dfrac{n_{k_{-i},C}}{M + n_T + n_C - 1} \int F_{\mathcal{N}}(X_{i,C}|\theta^*_{k,C})p(\theta^*_{k,C}|X_{k_{-i},C}, \theta_0)\mathrm{d}\theta^*_{k,C} \\ \dfrac{M}{M + n_T + n_C - 1} \int F_{\mathcal{N}}(X_{i,C}|\theta^*_{C})\mathrm{d}G(\theta^*_{C}|\theta_0) \end{cases}$$

Step 5. Sampling the mixture parameter θ^*_k when including the control group
For each component $k = 1, ..., K_T$, θ^*_k is set to be equal to $\theta^*_{k,T}$ (obtained in Step 2). And for each component $k = K_T + 1, ..., K$ which contains only the control group, the parameters $\theta^*_{k,C}$ has following conditional posterior:

$$\theta^*_{k,C}|X_C, \theta_0 \sim \prod_{i \in k} F_{\mathcal{N}}(X_{i,C}|\theta^*_{k,C}, \theta_0)G(\theta^*_{k,T}|\theta_0)$$

Step 6. Sampling the Gaussian latent variable when including the control group
As discussed in Section 2.3.3, $\{X_{i,t}, i = 1, ..., n_T\}$ of the treatment group are fixed at the values sampled in Step 3. Hence I only update $X^{(l)}_{i,C}$ given the full sample. For each *discrete* marginal $X^{(l)}$ out of *all* $X = (X^{(l)}, X^{(-l)})$ and for each component $k = 1, ..., K_T, K_T + 1, ..., K$, compute $\Sigma^{(l)(-l)}_k$ and $\mu^{(l)(-l)}_k$ using formulas (2.3.14), and sample $X^{(l)}_C$ from following truncated normal:

$$X^{(l)}_C|\theta^*_k, X^{(l)D}_k, X^{(-l)}_k \quad \sim \quad \prod_{i \in k} \Pr(X^{(l)D}_{i,C} = x|\theta^*_k, X^{(-l)}_{i,C}, X^{(-l)}_{i,C})\mathcal{N}(X^{(l)}_{i,C}|\mu^{(l)(-l)}_k, \Sigma^{(l)(-l)}_k)$$

2.4.2 PRIOR SPECIFICATIONS

Simply fixing the prior parameters at certain values might be dogmatic. To allow the model to be more flexible, I adopt hierarchical priors on the model parameters. Firstly, to assign a prior for the data with different scales, all marginals of X need to be standardized. I normalize continuous marginals of X into the $\mathcal{N}(0,1)$ scale by subtracting its sample mean and dividing by the sample standard deviation. Secondly, the base distribution $G(\theta_0)$ of the DP appears in its role as the prior distribution of normal mixture parameters $\theta_k^* = (\mu_k, \Sigma_k)$. I choose a normal-inverse-Wishart prior, i.e.:

$$
\begin{aligned}
\mu_k|\Sigma_k &\sim \mathcal{N}(\mu_0, \kappa_0^{-1}\Sigma_k) \\
\Sigma_k &\sim \mathcal{IW}_{\nu_0}(\Lambda_0)
\end{aligned}
$$

where ν_0 and Λ_0 are the degrees of freedom and the scale matrix of the inverse-Wishart distribution. Further, I assign the hyper prior on the variable $\theta_0 = (\mu_0, \Sigma_0, \nu_0, \kappa_0, \Lambda_0)$ and include Step 7 into the Gibbs cycle.

Step 7. Sampling prior on mixture parameters

This paper provides following suggestions on sampling the prior parameters $(\mu_0, \Sigma_0, \nu_0, \kappa_0, \Lambda_0)$:

1. μ_0 is set to be 0, which simply reflects the sample mean of the normalized variable.

2. Λ_0 is the scale matrix of the prior on the variance. Its diagonal elements are independently generated from: $\text{diag}(\Lambda_0) \sim \mathsf{Uniform}[0.2, 2]$. This is a relative tight prior, which can ensure that the shape of X can be approximated by enough mixture components.

3. ν_0, which can be thought of as a confidence of the prior, is sampled from: $\nu_0 \sim \mathsf{Uniform}[5, 15] + P + 1$, where P is the number of confounding variables.

4. κ_0^{-1}, which serves as a multiplier on the variance, is set to be: $\kappa_0^{-1} = 10 \times \nu_0$. This allows for a suitably wide range of possible locations of μ_k.

Finally, recall that M is a concentration parameter of the Dirichlet process. The concentration parameter M controls the number of clusters in a direct manner, with larger M implying a larger number of clusters a priori (Teh, 2011). To let the M be flexible, I adopt the approach by Jara et al. (2007) to assign a Gamma prior on the M. Therefore, Step 8 can be described as follows:

Step 8. Sampling the concentration parameter M

The concentration parameter M is sampled from: $M|a_0, b_0 \sim \mathsf{Gamma}(a_0, b_0)$ where $a_0 = 1, b_0 = 1$.

Nevertheless, based on my attempts and the results in Jara et al. (2007), the choice of (a_0, b_0) will not affect the resulting number of components significantly. In addition, Rossi (2013) also shows that the density estimates by DPM of normals are not strongly influenced by the value of the M parameter.

2.4.3 ATT Estimator

After obtaining MCMC sample $\{S^{(j)}, j = 1, ..., J\}$ from its target distribution by above Gibbs steps 1-8, I compute the ATT estimator. Based on posterior mean of the ATE (recall τ_h in Equation 2.2.3) in a subpopulation of matched set, the posterior mean of the ATT ($\mathrm{E}(\tau|Y, \mathcal{S}, X, \mathcal{H})$) and the the posterior variance of the ATT ($\mathrm{Var}(\tau|Y, \mathcal{S}, X, \mathcal{H})$) have closed forms as given in Equation 2.2.5 and 2.2.6, respectively. Thus, the Rao-Blackwell estimator of ATT τ is given by:

$$\widehat{\mathrm{E}}(\tau|Y, X, \mathcal{H}) = \frac{1}{J}\sum_{j=1}^{J} \mathrm{E}(\tau|Y, \mathcal{S}^{(j)}, X, \mathcal{H}) \tag{2.4.1}$$

$$\begin{aligned}\widehat{\mathrm{Var}}(\tau|Y, X, \mathcal{H}) &= \frac{1}{J}\sum_{j=1}^{J} \left[\mathrm{Var}(\tau|Y, \mathcal{S}^{(j)}, X, \mathcal{H})\right] \\ &\quad + \frac{1}{J}\sum_{j=1}^{J} \left[\mathrm{E}(\tau|Y, \mathcal{S}^{(j)}, X, \mathcal{H}) - \widehat{\mathrm{E}}(\tau|Y, X, \mathcal{H})\right]^2\end{aligned} \tag{2.4.2}$$

2.5 Data Applications

To validate the proposed DP matching method, I employ both the artificial data and the real data. Appendix B.1 to this paper provides two applications to the artificial data and compare the DP matching to various propensity score matching methods. In these two exercises, the EPBR conditions are satisfied because covariates are not ellipsoidally distributed. The first exercise shows that the DP matching can explicitly distinguish the mixture structure of the data. The estimates of the ATT via DP matching or propensity score matching are all close to the actual effect (Figures B.1-B.2 and Table B.1). The second application uses a synthetic data that the overlap of covariates in treated units and controlled units is relatively tiny. In the presence of strong separation between the treated and the controlled, the common support area of the propensity score becomes fairly small. Then as expected, the point estimates of the ATT via various propensity score matching methods are more biased, while the DP matching estimator can still successfully prune the 'noisy' controlled units and finally restore the benchmark value (Figures B.3-B.4 and Table B.2).

2.5.1 The Lalonde (1986) Data

Background

In real data application, this paper applies the well-known LaLonde (1986) data. With regard to observational studies, the LaLonde (1986) data has been commonly employed as a canonical benchmark for various causal inference approaches.[14] The data originated from a randomized experiment about a job training program in the US—the National Supported Work Demonstration (NSW).

Researchers (Dehejia and Wahba, 1999, 2002; Smith and Todd, 2005) re-analyze this data and replace the experimental control group by several other data sources. I restrict the application to the so-called NSW-PSID-1 subsample and NSW-CPS-3 subsample. The NSW-PSID-1 consists of 185 treated units from the original experimental data and 2490

[14] A detailed review of this dataset is provided by Li (2013), for example.

controlled units from the PSID data. The latter one includes a control group of 429 observations. In the data, the outcome of interest related to `earnings in 1978`. Pre-program individual characteristics include `earnings in 1974, earnings in 1975, age, years of education, Black` (dummy), `Hispanic` (dummy), `married` (dummy), `diploma` (dummy), `unemployed in 1974` (dummy), and `unemployed in 1975` (dummy). The counterpart of randomized program data provides a benchmark estimates of the ATE: the difference in sample means between the treatment group and the control group yields a treatment effect on post-program earnings of \$1794 with a 95 percent confidence interval of $[551, 3038]$.

DP Matching: a graphic analysis

First of all, to illustrate the mechanism behind the DP matching, I match on two variables using the NSW-PSID-1 subsample, i.e., `earnings in 1974` and `earnings in 1975`, and cross-plot their distributions using a typical MCMC iteration (Figures 2.2-2.3). Figure 2.2 plots all matched treated and controlled units. The treated units (marked as solid dots in the plot) have been well-approximated by 9 mixture components, which are highlighted in 9 different colors in the graph. The matched controlled individuals are marked as crosses in the plot. By the DP matching algorithm, they have been matched to one of the treatment group's component (referred to using the same colors in the graph). We can see from the graph that the matched controls appear to in similar covariates space as the treated. I then re-plot this graph by introducing the unmatched controlled units (Figure 2.3). The circles in navy color indicate unmatched controlled individuals. We can see that the majority of the unmatched controlled individuals have much higher earnings in 1974 and/or 1975, compared to those of the treated units and the matched controlled units. As suggested by Figure 2.3, the proposed DP matching method has successfully distinguished and then pruned these controlled individuals.

It is also very interesting to observe a heterogeneous pattern in earnings from Figure 2.2. The matched units in red can be viewed as people who were unemployed or who received very low annual income (less than about \$1000) in both years. Individuals plotted in green belong to the lower income group, whilst those in pink represent the middle class.

The matched pairs in yellow, orange and grey respectively indicate unemployed people or people with low income in either 1974 or 1975, but receiving higher income in the other year. The blue and grey colors indicate people from the higher income group.

Results: ATT estimators

Next, I estimate the ATT using the DP matching, and compare results with those generated by two nearest-neighbor matching methods (i.e.via the Mahalanobis distance and via the propensity score). Tables 2.1-2.2 summarize ATT estimates using various sets of confounding covariates in the NSW-PSID-1 subsample and the NSW-CPS-3 subsample, respectively.

Looking firstly at the NSW-PSID-1 subsample (Table 2.1), we find that, when matching on two, five or eight covariates (Panels A-C), the point estimates of the ATT by the DP matching are closer to the benchmark effect of $1794 than the ones by the nearest-neighbor matching via the Mahalanobis distance metrics. Despite the fact that the ATT estimates based on the Mahalanobis distance are less biased in the case of the ten matching variables (Panel D), the Mahalanobis distance metrics are pretty sensitive to different specifications with confounding variables. The signs of ATT estimates based on the Mahalanobis distance are positive in Panels A and D, but become negative in Panels B and C. The ATT estimates by the DP matching have the best performance in terms of biasedness when compared to other nearest-neighbor matching methods, when matching on two or five covariates. The propensity score matching estimates are less biased when eight confounding variables are included. However, they tend to deviate more from the benchmark effect than the ATT estimates by the DP matching once ten matching variables are used (Panel D of Table 2.1).

Table 2.2 tells a very similar story in the NSW-CPS-3 subsample. The DP matching method is still less sensitive to specifications with different matching covariates. When matching on two or ten covariates, the estimated ATTs by the DP matching are less biased than those generated by the Mahalanobis distance or the propensity score. It is interesting to note that when ten confounding covariates are used in the NSW-CPS-3 subsample, the nearest-neighbor matching methods will induce (Panel D of Table 2.2) negative ATT estimates, whereby the DP matching is still able to deliver an ATT estimate that is positive and closer to the benchmark effect. Furthermore, in both subsamples, the standard errors of all ATT estimates based on the nearest-neighbor matchings are consistently larger than the posterior standard deviations of ATT estimates using the DP matching.

Taken as a whole, it is safe to conclude that the DP matching method is robust to different specifications with different sets of confounding variables. Additionally, it can generate a well-performed ATT estimator in terms of biasedness and posterior standard deviation.

2.6 Conclusion

The proposed DP matching method in this paper is committed to delivering a desirable ATT estimator. It directly matches the controlled units to the treated group on their covariates space. Hence the balancing property can be met naturally. The joint distribution of covariates are approximated by Dirichlet process mixture of normals and the discrete confounders are augmented by Gaussian latent variables. The MCMC sampler integrates the matching procedure and the estimation of ATT into a single scheme. It takes matching uncertainties into account, and therefore, yields a credible confidence interval. When sampling from the posterior densities, hierarchical priors allow further flexibility of the model and the conjugate prior specification greatly simplifies the computation. The empirical application suggests that the ATT estimator via DP matching is competitive.

Additionally, it is always important to learn the potential limitations of an proposed approach. First, unlike the conventional frequentist approaches, the choice of prior specifications in the proposed approach might result in moderately different posterior ATT. Second, each covariate should contribute unequally to the matching procedure. For instance, the pre-treatment wage variables in the LaLonde (1986) data seem to have higher prediction power of the treatment assignments, and as a consequence should have greater contributions to the matching. The default prior specification suggested in Section 2.4.2

did not particularly address this issue. However, these limitations can be possibly avoided by choosing appropriate prior distributions. The exploration of the prior settings is left for the future work.

FIGURES AND TABLES

Figure 2.1: ILLUSTRATION OF THE PROBABILITIES OF MATCHING USING A SINGLE DIMENSIONAL EXAMPLE

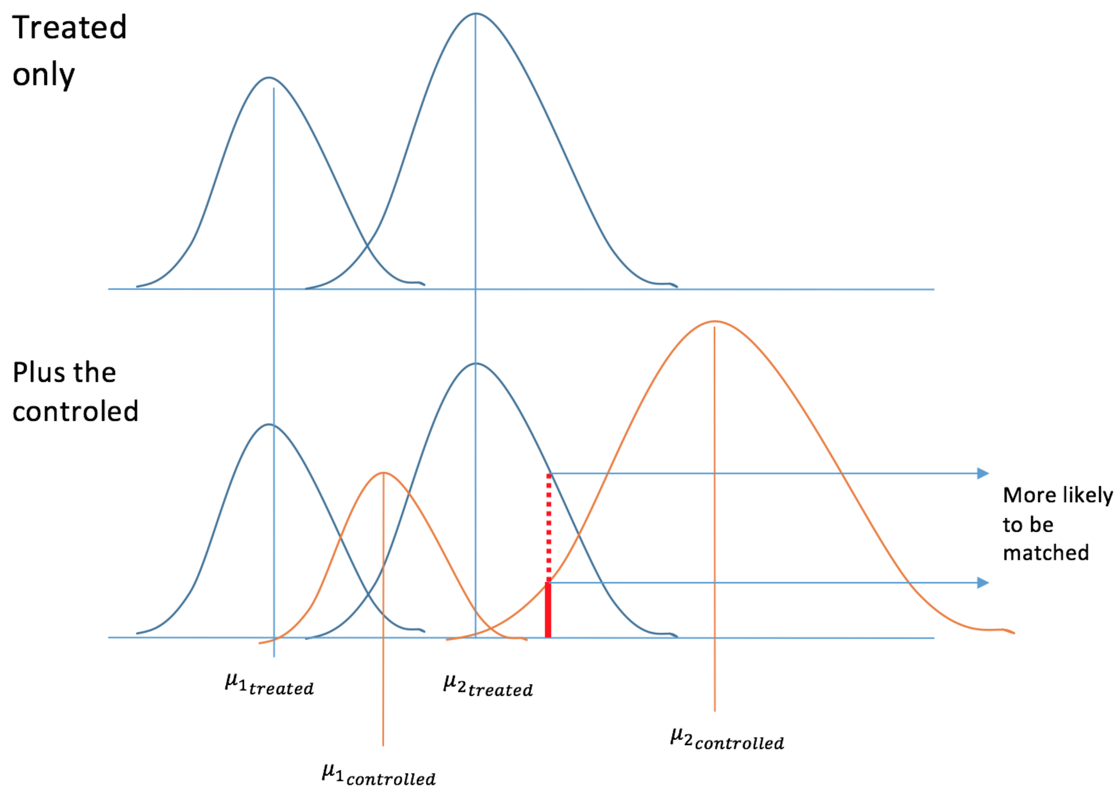


Figure 2.2: CROSS-PLOT OF MATCHED UNITS OF PRE-PROGRAM EARNING VARIABLES: 400^{th} MCMC ITERATION (NSW-PSID-1)

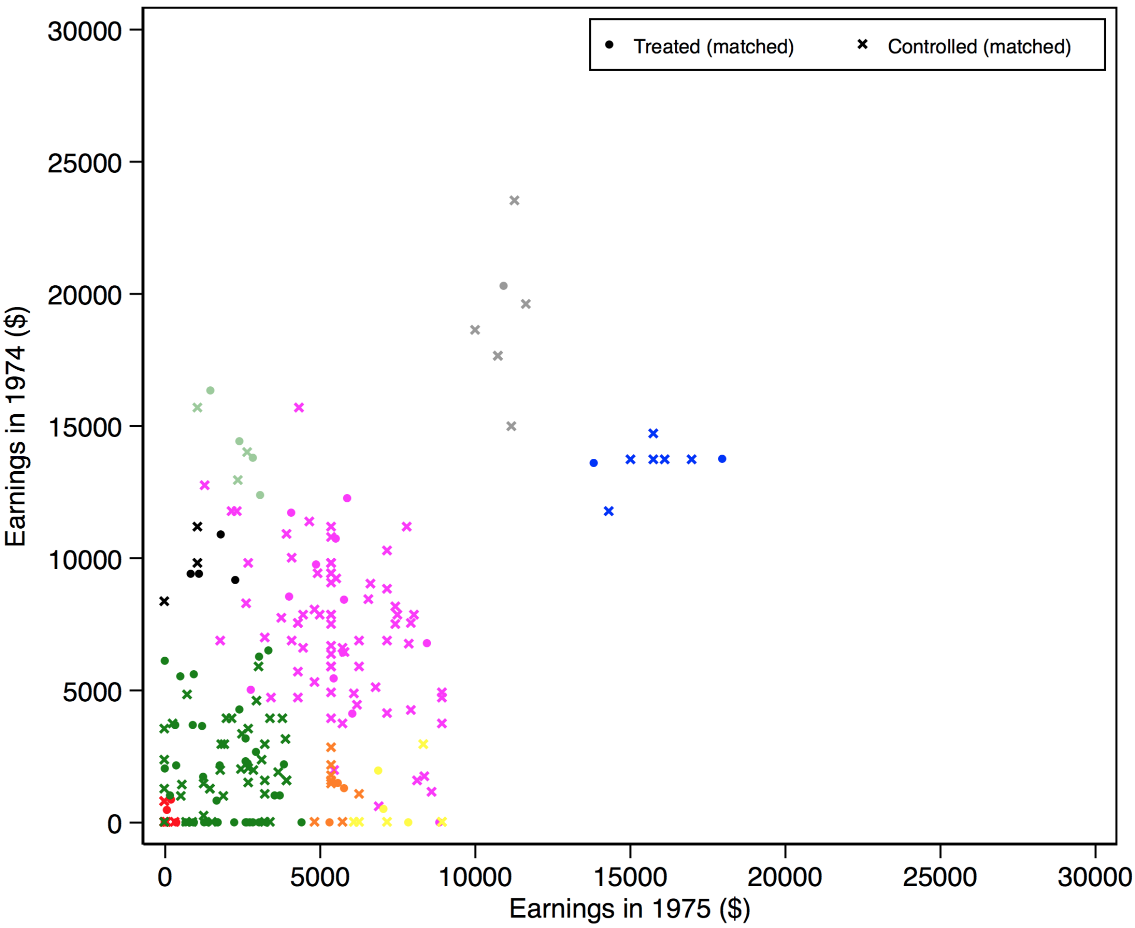


Figure 2.3: Cross-plot of Unmatched Units of Pre-program Earning Variables: 400^{th} MCMC Iteration (NSW-PSID-1)

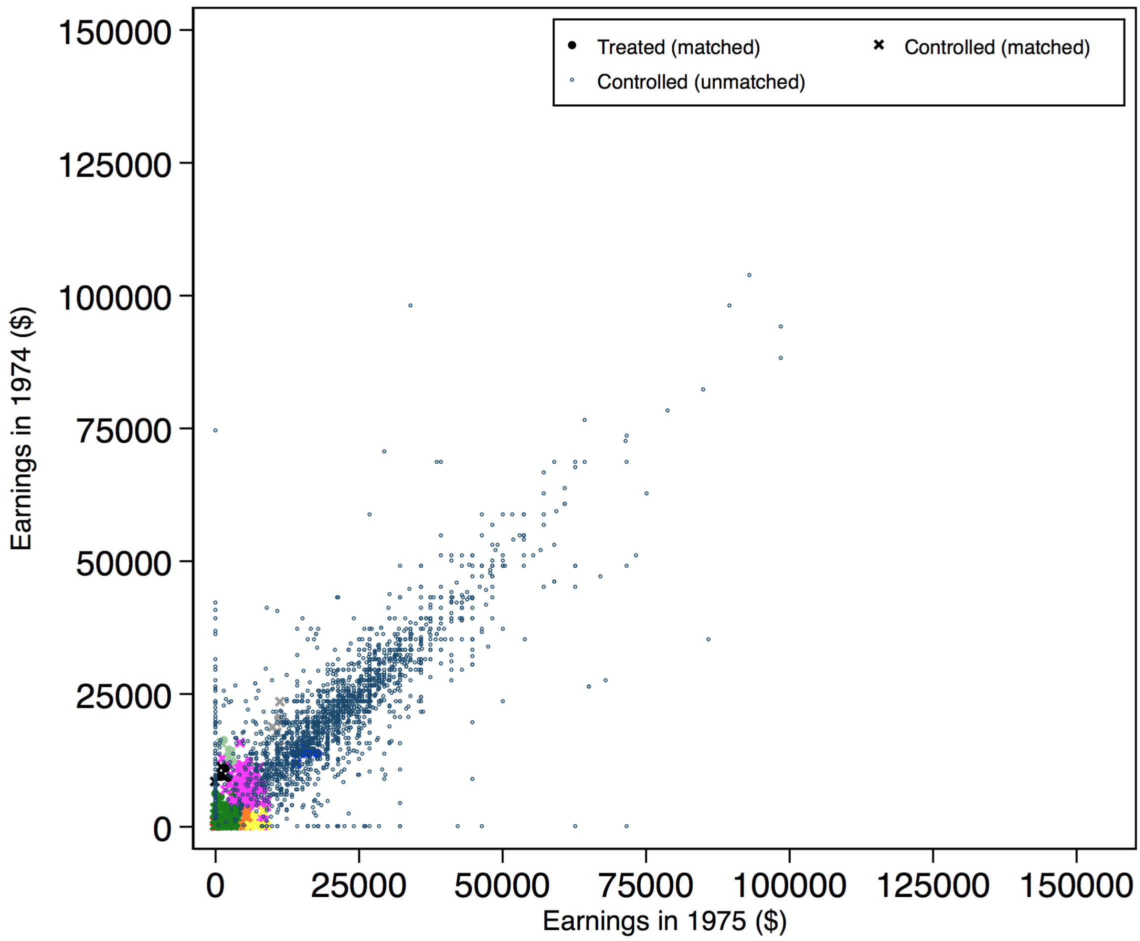


Table 2.1: Various Estimates for the Lalonde data (NSW-PSID-1)

	Est. ATT	SD/SE
Panel A: matching on two covariates		
DP matching	1810.928	(745.031)
Nearest-neighbor matching (Mahalanobis)	1754.493	(805.123)
Nearest-neighbor matching (propensity score-probit)	1436.375	(995.363)
Nearest-neighbor matching (propensity score-logit)	1706.182	(992.125)
Panel B: matching on five covariates		
DP matching	1620.702	(885.757)
Nearest-neighbor matching (Mahalanobis)	−273.007	(1214.134)
Nearest-neighbor matching (propensity score-probit)	890.535	(1144.828)
Nearest-neighbor matching (propensity score-logit)	1511.544	(1074.692)
Panel C: matching on eight covariates		
DP matching	2258.507	(954.075)
Nearest-neighbor matching (Mahalanobis)	−739.543	(1093.066)
Nearest-neighbor matching (propensity score-probit)	1545.521	(1129.796)
Nearest-neighbor matching (propensity score-logit)	2125.715	(1004.261)
Panel D: matching on ten covariates		
DP matching	2442.980	(815.443)
Nearest-neighbor matching (Mahalanobis)	2050.486	(924.825)
Nearest-neighbor matching (propensity score-probit)	2660.667	(1252.966)
Nearest-neighbor matching (propensity score-logit)	925.816	(1684.988)

Notes: First row in each panel presents MCMC posterior mean and standard deviation based on DP matching. Bootstrap with 100 replications is used to estimate standard errors for propensity score matching. The benchmark experimental point estimate=1794 with 95% CI$\in$ $[551, 3038]$.
Panel A is based on the matching on `earnings in 1974` and `earnings in 1975`.
Panel B is based on the matching on `earnings in 1974`, `earnings in 1975`, `age`, `years of education` and `black`^{D}.
Panel C is based on the matching on `earnings in 1974`, `earnings in 1975`, `age`, `years of education`, `black`^{D}, `hispanic`^{D}, `married`^{D} and `diploma`^{D}.

Panel D is based on the matching on `earnings in 1974`, `earnings in 1975`, `age`, `years of education`, `black`^{D}, `hispanic`^{D}, `married`^{D}, `diploma`^{D}, `unemployed in 1974`^{D}, and `unemployed in 1975`^{D}.

Table 2.2: Various Estimates for the Lalonde data (NSW-CPS-3)

	Est. ATT	SD/SE
Panel A: matching on two covariates		
DP matching	869.264	(726.635)
Nearest-neighbor matching (Mahalanobis)	738.679	(832.7141)
Nearest-neighbor matching (propensity score-probit)	467.760	(924.903)
Nearest-neighbor matching (propensity score-logit)	257.680	(738.636)
Panel B: matching on five covariates		
DP matching	1161.169	(879.794)
Nearest-neighbor matching (Mahalanobis)	1028.723	(1221.974)
Nearest-neighbor matching (propensity score-probit)	1918.446	(1492.876)
Nearest-neighbor matching (propensity score-logit)	867.269	(1037.917)
Panel C: matching on eight covariates		
DP matching	1364.988	(860.398)
Nearest-neighbor matching (Mahalanobis)	262.240	(1207.09)
Nearest-neighbor matching (propensity score-probit)	1968.800	(1179.098)
Nearest-neighbor matching (propensity score-logit)	2125.715	(1004.261)
Panel D: matching on ten covariates		
DP matching	1353.728	(858.495)
Nearest-neighbor matching (Mahalanobis)	−415.500	(1203.15)
Nearest-neighbor matching (propensity score-probit)	−1252.038	(1284.344)
Nearest-neighbor matching (propensity score-logit)	−1042.433	(1380.570)

Notes: First row in each panel presents MCMC posterior mean and standard deviation based on DP matching. Bootstrap with 100 replications is used to estimate standard errors for propensity score matching. The benchmark experimental point estimate=1794 with 95% CI$\in [551, 3038]$.

Panel A is based on the matching on `earnings in 1974` and `earnings in 1975`.

Panel B is based on the matching on `earnings in 1974`, `earnings in 1975`, `age`, `years of education` and `black`^{D}.

Panel C is based on the matching on `earnings in 1974`, `earnings in 1975`, `age`, `years of education`, `black`^{D}, `hispanic`^{D}, `married`^{D} and `diploma`^{D}.

Panel D is based on the matching on `earnings in 1974`, `earnings in 1975`, `age`, `years of education`, `black`^{D}, `hispanic`^{D}, `married`^{D}, `diploma`^{D}, `unemployed in 1974`^{D}, and `unemployed in 1975`^{D}.

Chapter 3

Empowering Mothers and Enhancing Early Childhood Investment: Evidence from a Unique Preschool Program in Ecuador [1]

3.1 Introduction

Empowering women and enhancing early childhood development are important policy goals which are often pursued as separate, discrete initiatives in developing countries. We study an approach that exploits potential complementarity by pursuing both goals through one measure, based on empowering mothers and teaching parenting skills that advance their children's development. Empowerment for mothers in this program is based on acquiring knowledge, home and personal practices that strengthen their role in the family and community and increase awareness of their role in their children's education. These sessions

[1] We thank the AVSI staff in Quito, particularly Sarah Holtz, Amparito Espinoza, Stefania Famlonga, Lucy Troya, Delia Llanos and Veronica Echeverría and the AVSI staff in Milan, particularly Maria Teresa Gatti, Andrea Bianchessi and Alberto Piatti for their collaboration and logistical support through the design, field work and data-collection phases of this study. We thank Sascha O. Becker, Gabriella Conti, Clément de Chaisemartin, Rocco Macchiavello, Fabian Waldinger, Christopher Woodruff, and participants at CAGE seminar in the Department of Economics at the University of Warwick and at the RIDGE/NIP-LACEA Workshop on Inequality, Poverty and Politics for useful comments and suggestions. We thank CAGE and the Department of Economics at the University of Warwick for financial support for this project. Victor Lavy acknowledges financial support from the European Research Council through ERC Advance Grant 323439.

include joint activities for mothers and their children, and separate group activities for the children in the program.

The accumulation of evidence that education programs are both more effective and less costly when delivered to younger children has sharpened the focus on early childhood research and policy (Heckman et al., 2010, 2013; Gertler et al., 2014). However, questions remain about which policy tools are most effective.

Similarly, empowering women has been the focus of research and policy because engaging women as equal participants in the community and economy has been seen to enhance development outcomes (Klugman et al., 2014; Wong, 2012). Doing both - empowering women in a way that also supports early childhood development - has not been widely studied, however, it could be a means of achieving both goals. This approach has the potential to deliver improvements in women's empowerment and status at home and in the community, as well as improving children's education. Duflo (2012) defines women's empowerment as "improving the ability of women to access the constituents of development – in particular health, education, earning opportunities, rights, and political participation". Improving mothers' access to each of these domains can have a positive effect on early childhood, which has been shown to deliver life-long benefits (Barber and Gertler, 2009; Carneiro et al., 2013; Kiernan and Huerta, 2008)[2]. Evidence of spillover effects from early child development programs on the empowerment of their mothers is scarce and needed (Baker-Henningham and López Bóo, 2010).

In this paper we study the consequences of a home-preschool program designed to enhance both women's empowerment and children's early childhood development. The *PelCa* (preescolar en la casa – home pre-schooling) program started in Pisullì, one of the poorest neighborhoods of Quito, Ecuador, in 2005. It is run by the Association of Volunteers in International Service (AVSI), an international non-governmental organization

[2] Carneiro et al. (2013) find that maternal education leads to more positive home environments, particularly for low-ability mothers. Moreover, mothers with more schooling invest more quality time with their children (breastfeeding, reading to them, taking them out) and this is later reflected in better cognitive outcomes and fewer behavioral problems. According to Barber and Gertler (2009), empowering women to become informed and active health consumers reduces child mortality, morbidity, anemia, and stunting. Kiernan and Huerta (2008) show that economic deprivation and maternal depression negatively affect children's cognitive and emotional outcomes, partly because of less nurturing and parental engagement.

founded in Italy that focuses on human development. The program is open to mothers with children aged up to three years old, and currently involves hundreds of mothers and children.

In the program we studied, a qualified family advisor trained groups of six to eight mothers. Children accompanied their mothers in these group sessions, which were held every two weeks in the NGO offices. There were three parts to each group session. In the first part, mothers received structured training focused on strengthening their role in the family, and learned parenting techniques that emphasized their children's early development. At the same time, children socialized using educational games and didactic materials. In the second part, family advisors taught mothers and children educational activities that could be reproduced at home, to improve the quality of maternal-child interactions and enhance mothers' participation in child development activities. In the third part, advisors monitored and assessed the home assignments from the previous meeting.

The program was implemented non-experimentally, but since its initiation in 2005, new families have joined every year. Assuming that new applicants resemble those who joined the program earlier, we selected a control group from the applicants in 2012. Following our suggestion, the NGO made a special effort to reach as many eligible families as possible in 2012. This provided us with a large pool of applicants from which we selected our control group, which consisted of families that had an older child in any grade in primary school and a younger child who would enroll in the program jointly with his or her mother. This can be viewed as a quasi-natural experiment, and we will demonstrate that it yields well-balanced treatment and control groups.

We evaluate the effect on women's empowerment after two to seven years of participation in the program by focusing on: self-esteem, inputs into child education, labor-market participation and earnings, allocation of decision-making within the household, and economic and social independence. We also examine the impact on children's educational outcomes, such as how likely children were to repeat a grade or drop out of school, and how they fared in cognitive tests. Our evidence shows that the program empowered women in various dimensions: mothers who participated in the program for two to seven years are more likely to be employed, more likely to have a full-time job, and more likely to have a formal-sector job. Mothers who have been in the program also earn higher wages, and are more likely to manage their own money and to make independent decisions about how to spend it. Women's autonomy is also reflected in a higher likelihood of deciding by themselves whether to work outside the home, and a higher likelihood of returning to school as an adult. Moreover, there is evidence that these women take on a greater role overall in intra-household decisions, especially on matters involving children's education and discipline. Mothers who participated in the program increased their child investment, for example by spending more time practicing cognitive and social skills with their children.

The program had mixed effects on children: it significantly reduced the drop-out rate and likelihood of temporarily withdrawing from school, and it improved scores in cognitive tests (though it is precisely measured only for some sub-groups of the overall sample). However, we find no effect on children's self-reported or mother-reported attitudes towards schooling.

Allowing for heterogeneity of treatment effect by mother's pre-program characteristics and outcomes and by child gender reveals meaningful differences, which we use to interpret our findings. All of the above results hold when we estimate aggregate treatment impacts, using summary indices instead of individual outcomes, in order to account for multiple inference, when we use entropy balancing to adjust for differences in pre-treatment covariates, and when we use other robustness checks.

The remainder of our paper is structured as follows. In Section 3.2 we present an overview of the literature on women's empowerment and on early childhood development. Section 3.3 outlines the background and design of the quasi-natural experiment. In Section 3.4 we describe the data and in Section 3.5 we discuss the empirical analysis and results. In Section 3.6 we explore potential mechanisms through which results are achieved and discuss robustness checks. The conclusion is in Section 3.7.

3.2 Related Literature on Early Childhood Interventions

The present work is related to two different literatures: studies on women's empowerment and studies on early childhood development. The literature on women's empowerment is more extensive. Kabeer (2005) defines empowerment as the "ability to make choices" in ways that change power relations and affect women's education, employment, and political participation. Duflo (2012) defines women's empowerment as improving women access and utilization of social, political and economic opportunities. Decision-making within the household is also an important indicator of the distribution of power within the household (Alkire, 2007; Narayan-Parker, 2005). We will follow this approach and explore intra-household decisions, capturing women's power relations within the household and their access to the constituents of development (e.g. whether they are allowed to work). Different channels for empowering women have been explored. Education is sometimes proposed as one of the main drivers of empowerment (Oyitso and Olomukoro, 2012) , but the evidence is mixed. There is substantial evidence that education can improve cognitive skills, raise aspirations, allow access to information, raise awareness to real conditions, and help coping with dis-equilibrium (Kabeer, 2005; LeVine et al., 2001). More educated women also experience less domestic violence (Kabeer, 2005; Sen, 1999). Mocan and Cannonier (2012) find that more educated women in Sierra Leone are "more intolerant of practices that conflict with their well-being". It is less clear, however, whether this change in preferences translates into behavior. Andrabi et al. (2012) show that higher maternal education improves maternal child care, but the study does not find an effect on intra-household decision-making.

Women are also empowered by accumulating wealth but there is little evidence on how to enhance the mechanism. Microfinance programs can facilitate the accumulation of economic assets but evidence for a causal effect of such programs on women's empowerment is mixed. Kabeer (2001, 2005) suggests that women's access to credit improves women's self-perception, reduces domestic violence and increases women's power in the household decision-making process. In households where the loan recipient was male, the role of women in decision-making regarding loan use, enterprise management and allocation of profits was much lower than in households where the loan recipient was a female. However, Banerji et al. (2013) find no short- or long-run effects of micro-credit in India on women's empowerment.

Baker-Henningham and López Bóo (2010) suggest that training mothers in group parenting sessions is a cost-effective method of service delivery but little is known about the effect of such programs on mothers' longer term well-being and life course outcomes. Our study is the first to focus on long-term exposure to group parenting sessions and its impact on both mothers and their children and it contributes to the growing literature on early age interventions. Among others, we note the Abecedarian project, High Scope Perry Preschool Program, Chicago Child-Parent Centers and the Head Start Program. Evidence suggests that these programs led to improved schooling attainment and better outcomes in adulthood (higher employment rate and earnings, lower crime rates). For example, the pre-school Abecedarian project improved children's reading and mathematics achievements, lowered grade retention and increased completed education at adulthood (Campbell et al., 2002; Temple and Reynolds, 2007). The High Scope Perry preschool program affected the schooling outcomes of girls only: by age 19 treated females had a higher school GPA and completed a higher grade (Heckman et al., 2013); by age 27 treated females were 30% less likely to drop-out from high-school (Nores et al., 2005). We note however, that these interventions target the most-disadvantaged groups and that such programs may be unfeasible in most developing countries because they are expensive. Most related evidence in developing countries is often based on very short interventions and small samples (see Baker-Henningham and López Bóo (2010); Nores and Barnett (2010) for a literature review). Few studies focus on longer treatment and long-term child outcomes. Exceptions are Watanabe et al. (2005) and Kagitcibasi et al. (2009), both providing evidence of positive effects on cognitive outcomes, while Kagitcibasi et al. (2009) find positive effects on other socio-economic outcomes. For example, children exposed to an early treatment entered the workforce later (due to longer schooling) and found jobs of a higher status as young adults.

Another related study is Rosero and Oosterbeek (2011), which evaluates the effect of home visits and child care centers on mothers and children in Ecuador. Child care centers provide day care for the whole day throughout the entire year, and groups of 8-10 children are supervised by a trained teacher. Weekly home visits, each lasting about an hour, teach mothers how to stimulate and nourish their children, in individual sessions if children are younger than three, and in groups if children are older. Rosero and Oosterbeek (2011) find that child care centers increase mothers' labor force participation but have detrimental effects on children's cognitive and health outcomes, while they observe the opposite effects for home visits. The early childhood development program that we study is more similar to the home visits than to the child centers described above.[3] However, while Rosero and Oosterbeek (2011) can analyze short term treatment effects only[4], we are able to explore long term effects on the mother and children. Attanasio et al. (2014) also evaluate the effects of a weekly home visit program in Colombia that targeted children 12-24 months and lasted for 18 months. They find that enhancing mothers' engagement with their children positively affects their cognitive and socio-emotional domains through an increase in parental investments and find no effects on mothers' depression (Attanasio et al., 2014, 2015).

3.3 Background and Research Design

AVSI, the Association of Volunteers in International Service, is an international not-for-profit, non-governmental organization (NGO) based in Milan, Italy. Founded in 1972, it operates in 30 countries in Eastern Europe, Africa, Latin America and the Middle East and it runs more than 100 long-term projects. It started its activities in Ecuador in 2001, focusing on infant and child development and education. In 2005 an AVSI branch

[3] The main difference between the *PelCa* program and the home visits analyzed by Rosero and Oosterbeek (2011) is that in the latter mothers with their children are visited in their homes once a week, and individually if the child is younger than 3, while in the *PelCa* program mothers and children go to the NGO twice a month and are taught in groups from the beginning.

[4] Data was collected after the children in the program were exposed to treatment during 21 months (Rosero and Oosterbeek, 2011).

was opened in Pisullí, a disadvantaged, urban neighborhood to the northwest of Quito. In collaboration with Fundación Sembrar, a local non-profit organization, and the local parish, AVSI funded a community development center where it implements a modified version of *PelCa*. The program expanded rapidly, providing after-school programs and other services to more than 700 children, youth and their families in 2013, with more than 50 members on the local staff.

3.3.1 The Intervention

PelCa is a preschool program targeted at parents of children age 5 and under, based on group-parenting sessions. Fortnightly meetings are held in the NGO for small groups (usually six-eight mothers – with their children), under the guidance of a family advisor. In the first part of the meeting, children socialize with each other, playing games using didactic materials, while parents read and discuss material about family education. In the second part of the meeting, parents and children work together: they learn songs, educational games and various development activities that parents can reproduce with their children at home (e.g., reading books, playing with puppets, playing building games, etc.). The family advisor gives every child a notebook of age-appropriate activities that focus on different areas of development, and parents and children are expected to undertake these activities in the two weeks between the program sessions. In the last part of the meeting, the family advisors verify whether tasks that were assigned to the mothers in the previous two weeks were completed. The family advisors then monitor the progress made by each child, verify whether they have completed home assignments with the parent (e.g., by having children show drawings, or having children answer questions based on a story that was to be read to them by the parent). The family advisor gives each parent and child reinforcement activities to perform at home in the next two weeks. These activities are geared towards mothers and children achieving specific targets and goals.[5] These NGO-set goals are the basis for the outcomes that we evaluate in the paper.

Goals for the mother: Build self-confidence and self-awareness and a greater ability to

[5]The goals are taken from the NGO handbook in Spanish.

relate to their environment and the people in their peer group; to build a sense of value in oneself and in material assets.

At the personal level: awaken interest in life; assume responsibility through personal commitment; enhance self-perception of own abilities and the capacity to take initiative; and to give appropriate value to assets and saving.

At the level of relationships with the family and community: to take pleasure in dialogue, patience and reflection; strengthen the level of involvement of each member of the family in the light of existing dynamics; share between the couple the responsibility for the children's education; develop an attitude that favors the autonomy of children and teenagers; develop relationships of intimacy and solidarity with the group in the meetings, and with neighbors in the neighborhood.

Goals for the children: facilitate children's integral growth in different areas of development (psychometric, language, cognitive, socio-affective).

Families usually find out about the program via word of mouth or a poster outside the NGO. If they express interest, AVSI employees visit the family at home to collect information about the family circumstances, observe conditions at home, assess the need for support, and identify family weaknesses and strengths. Children up to three years old are eligible to enter the program (so that they can participate in the program for at least two years). The mother commits to participating in fortnightly meetings and to performing the assigned tasks at home. The selection process also takes into account the family's financial standing and the proximity of the home to the NGO sites where sessions are held. Parents and children can remain in the program until the child is 12 years old but once the child reaches five years of age they move to the NGO *PelCa* school program. The application process for the *PelCa* pre-school program starts at the end of April and lasts for two weeks. Approximately 50 families (the number can vary depending on funding for that year) are selected to start the program in September.

3.3.2 Design: choosing a comparison group

Our treatment group was selected from many applicants, and is made up of mothers who enrolled in the *PelCa* preschool program and their children, who are now in primary school. In the summer of 2012, we selected a control group by mimicking the program's selection process, but on a larger scale. The NGO advertised the *PelCa* program in schools and through posters, as had been done previously in the program. However, it extended the application period to approximately two months to reach as many families as possible, and, indeed, attracted a much larger pool of applicants compared to other years. We selected applicant families with a preschool-age child and, similar to our treated mothers, at least one child enrolled in primary school. The identifying assumption is that the sample of mothers with children of primary school age who did not participate before in *PelCa* but chose to do so now with a younger child, represents a sound counterfactual for *PelCa* mothers and their primary-school-age children.

The families making up the control and treatment groups were invited to an interview in June-July 2012. The mother participated in a structured interview, while her primary school children were tested for cognitive and non-cognitive skills. The mother was asked to bring her children's vaccination certificate and birth certificate, which includes a record of the child's birth height, weight and head circumference and the older child's school report cards for the previous and current years. We then selected from among the applicants, those who had an older child at a primary school age. We also held a follow up interview with these control and treatment groups a year later in the summer of 2013.

3.4 Data

The data were collected through face-to-face interviews with mothers and children, using a questionnaire we developed specifically for this study.[6] The questionnaire seeks information

[6] We piloted the survey questionnaire in January 2012, interviewing 23 treated mothers: 12 of them had a primary-school-age child who participated in the *PelCa* preschool program and 11 of them had a primary-school-age child who did not participate in the *PelCa* program. We revised the questionnaire following this pilot test.

on family members (mother, partner and children), demographic characteristics, labor-market activities (type of job, full-time/part-time, formal/informal sector, wage, etc.), intra-household decision-making, and parents' inputs into child rearing. All questions targeted information current or prior to enrollment in the program.[7] The interview lasted approximately 45 minutes. The mother then took the Big Five Personality Test[8] and the Rosenberg self-esteem scale.[9]

Each child took cognitive tests in Spanish and mathematics.[10] Data on weight, height and head circumference at birth of the school age children were gathered through vaccination certificates and birth certificates. For some children, this information was incomplete or unavailable.

The control group for 2012 was 164 children and 115 mothers, while the treatment group was 219 children and 166 mothers, a total of 383 children and 281 mothers. We interviewed some grandmothers who participated in the program on behalf of the mothers, but we excluded these families from the analysis because we do not have grandmothers in the comparison group.

In summer 2013, we conducted follow-up interviews. Ten female interviewers from the area conducted home visits with all of the mothers in the sample. To obtain comparable information in the two rounds of data collection we used the same questionnaire, but with slight modifications. Some questions were added in order to clarify issues we encountered in the 2012 data. However, where we introduced new questions, we also sought retrospective information. Mothers were asked to bring the vaccination and birth certificates again (because many of these documents had been missing in the previous year). Eventually

[7]We will provide details about the questionnaire upon request.

[8]The Big Five Personality Test is based on decades of research. It consistently evaluates five broad traits of personality through a series of questions: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. In 1981 these factors became known as the "Big Five" to indicate the broad dimensions to which they refer. It has since been used intensely.

[9]The Rosenberg test, developed in 1965 by Dr. Morris Rosenberg, includes 10 Likert-type questions and it is used to evaluate the self-esteem of an individual and is widely used today by psychologists, sociologists, and social scientists. It has been translated into various languages (e.g., French, Norwegian, Spanish, Portuguese, Chinese, and Italian).

[10]Grade specific language and mathematics tests were constructed based on national tests from Ecuador, known as "pruebeas SER (Sistema de Evaluación y Rendición de la Educación)" and on Peruvian national tests for grades for which Ecuadorian tests were not available.

we collected data on height at birth for 44 percent of the children, weight at birth for 41 percent, and head circumference at birth for 38 percent. We think that the extent of missing values for these variables preclude a meaningful analysis of this information. Mothers were also asked to bring children's school report cards for 2010-11 and 2011-12.The children were tested again in Spanish and mathematics, using tests appropriate for the student's school grade. The follow-up sample included 136 control children (83 percent) and 98 control mothers (85 percent), and from the treatment group, 197 children (90 percent) and 150 mothers (90 percent).

3.4.1 Treatment-Control Comparisons: Balancing Tests

We examine in this section whether pre-treatment covariates are balanced between treatment and control groups. The evidence suggests that mothers and children in both groups are very similar on observed and predetermined characteristics, supporting our view of the empirical setup as a quasi-natural experiment. The first two columns of Tables 3.1-3.2 display the means for the treatment and control groups, while the last two columns present the difference in means between the two groups and the standard error.

Mothers' characteristics (Table 3.1) are balanced in most of the dimensions, except that control mothers were more likely to be employed before joining the program: 47.0 percent of treated mothers were working versus 60.9 percent of control mothers. Control mothers were also more likely to be working full-time. Pre-treatment paternal characteristics and most of the household characteristics are well balanced,[11] with a few exceptions: whether the family owned a house, the number of rooms and the availability of drinkable water in the house. Overall, 3 of the 33 pre-treatment maternal characteristics differences are significantly different at 10 percent level of significance. This is less than 10 percent. The F-test on the bottom of Table 3.1 on all of the pre-treatment maternal characteristics[12] is also not significantly different from zero, suggesting that the imbalance in the pre-program employment status of the mothers is an exception. We will however include in

[11] Details are provided in the Table C.1 and C.2 in the Appendix.
[12] Pre-treatment characteristics related to intra-household decisions before treatment are included.

the regressions pre-treatment control variables to capture these differences between the treatment and control groups. It will be shown that the estimates are not sensitive to adding these controls.

With respect to the child characteristics (Table 3.2), control group children are half a year older on average. This is probably due to the fact that we selected children from 1st to 7th grade and that, as we will see, control children are more likely to repeat a grade and therefore be older at survey date. The F-test on the significance of all the characteristics together suggests that overall children's characteristics are not linearly correlated with treatment status.

3.4.2 Entropy Balancing

An alternative way to control for the differences in some of the pre-treatment characteristics is to use entropy balancing (Hainmueller, 2011). Entropy balancing is a data-preprocessing method to achieve covariate balance. It computes the means (or higher moments of covariate distributions) of the covariates in the treatment group and looks for a set of entropy reweights so that the means in the reweighted control group match the means in the treatment group. We implement entropy balancing for the means of the covariates that we will include as control variables in our analysis (child and household pre-treatment economic characteristics).[13] Entropy balancing makes the treatment-control covariate balance almost perfect: differences in means are not significantly different from zero for all covariates (Table C.3).[14] This approach is preferred over a propensity score matching because the former eliminates all treatment-control imbalances. In addition, the propensity score matching requires treated and controlled units to be comparable within the common support. As a consequence, the individuals who do not lie in the common support (5 out of 281 mothers in 2012 and 22 out of 496 mothers in the pooled data) are dropped from the sample.[15]

[13]We also include an indicator of whether mothers were working full-time before treatment.

[14]We also obtain balanced samples through entropy balancing when we consider the children's samples or when we pool the two years of data together.

[15] For purposes of robustness, we also re-estimated the effects of the program through a propensity score matching, with and without replacement, and the results are very similar to the estimates we present in

3.5 Empirical Strategy and Results

We estimate the effect of participating in the program on the outcomes of interest using the following regression model:

$$\begin{aligned} y_{it} &= \beta_0 + \beta_1 Treatment_i + \beta_2 ChildCharacteristics_i \\ &\quad + \beta_3 HouseholdDemographics_i + \beta_4 HouseholdEconomics_i \\ &\quad + \beta_5 TimeFE_t + \beta_6 SchoolFE_i + \varepsilon_{it} \end{aligned} \tag{3.5.1}$$

where i is the individual and t is time. y_{it} is a vector of maternal and child outcomes of interest. Since we face a multiple outcomes problem, we will also compute summary indices[16,17] for domains of outcomes. $Treatment_i$ is a dummy equal to 1 when mother and child participate in the *PelCa* program and 0 otherwise. To shed light on heterogeneous treatment effects by number of years of participation in the program, we will also use the specification outlined above with a linear effect of number of years a mother/child participated in the program. Exposure to the program varies from two years and four months to seven years and eleven months.

$ChildCharacteristics_i$ includes year of birth, birth order, number of siblings as of 2005, i.e. before the program started, and gender. $HouseholdDemographics_i$ are pre-treatment household demographic characteristics: mother's and father's age, their civil status (married, lived together, mother was single) at the time of the birth of the first child, and the parents' level of education before the birth of the first child, a dummy equal to 1 if the mother was born in Quito, a dummy equal to 1 if the parents came from the same city, and the number of children the mother had in 2005. $HouseholdEconomics_i$

the paper (these results are available upon request).

[16]We followed Kling et al. (2007) to construct each summary index as an "equally weighted average of z-scores of its components, with the sign of each measure oriented [...] so that more beneficial outcomes have higher scores. The z-scores are calculated by subtracting the control group mean and dividing by the control group standard deviation."

[17]We developed a Stata package "**mseffect**" to calculate the mean effect size on the summary index with the advantage that we account for different weights, reversibility of outcome sign, and different types of robust standard errors.

are pre-treatment household economic characteristics: whether the mother worked before treatment, whether the father worked, the mean firm size of mother's and father's employer, average monthly family income before treatment. $TimeFE_t$ is a dummy equal to 1 when the observation corresponds to 2013, 0 if 2012; is the error term, clustered at the mother level when we run regressions pooling the observations in the two years together or when we analyze outcomes for children. $SchoolFE_i$ are school fixed effects.[18] They are included when we analyze outcomes for children. A more detailed description of the control variables is provided in the Appendix.[19]

3.5.1 Results based on Summary Indices

As we note in Section 3.3.1, the primary purposes of the *PelCa* program are to empower mothers, harmonize intra-family relations, and increase investment in education in early childhood. The breadth of the goals implies that the consequences of the program can be measured in multiple domains. We decide to measure the following domains, each of which contains multiple outcomes for mothers and children. For mothers, the domains (and specific outcomes) are: labor-market outcomes (whether working, working full-time, working with a contract and average family monthly income), mothers' economic and social independence (whether she has control over her own money, participates in voluntary activities, is currently studying and whether she has a role in her own employment decisions), mothers' intra-household decision-making (role in decisions regarding child's education, health, and discipline, expenditures in general and on food, having children, use of contraceptives, own health), mothers' child investment (own time inputs with child, aspirations/expectations for child's education), mothers' self-esteem and Big Five personality traits (Rosenberg scale, agreeableness, conscientiousness, extraversion, neuroticism and openness to experience). Summary statistics on the domains indices and individual outcomes are presented in Table 3.3, in panel A for mother's labor market outcomes, in panel B for mother's economic and social independence outcomes, in panel C for mother's role

[18] 55 are the number of schools that children attend in 2012.

[19] In a later version of this paper, we have tried different specifications with control variables.

in intra-household decision making, in panel D for mother's child's investment, in panel E for mother's non-cognitive skills and fertility choices, and in panel F for the father's labor market outcomes.

For children, the domains (and outcomes) are: test scores (language and mathematics test score), school dropout and grade repetition, attitude towards schooling (whether child likes school, whether child feels he is expected to follow certain behavioral rules and whether the child chose a book as gift[20]).

Before presenting the detailed estimates of the effects on each specific outcome, we analyze each domain by creating domain-specific summary indices. This allows us to control for the potential problem of over-rejection of the null hypothesis due to multiple inference. Because different outcomes have different data scales, simply averaging the estimators for the treatment effect is not likely to produce a meaningful statistic. To address this concern, we follow the summary-index approach per Kling et al. (2007). The summary index of multiple outcomes is the average of z-scores of each outcome variable. Z-scores are calculated by subtracting the control mean from the outcome and dividing by the control standard deviation. This summary index is a special case of the z-score[21] and is identical to the mean effect size of treatment if there is no missing value.[22] In general,

[20]At the end of each interview, children were offered a gift. They could choose between a book and a game. We interpret the choice of a book as interest in schooling activities.

[21]Here we replace the minuend and the divisor in the z-score by the control group mean and standard deviation respectively. In other words, we do require some dispersion in the controlled outcomes to guarantee the validity of standardization.

[22]In the regression specification this approach yields standardized estimators as follows: the treatment effects for K outcomes are aggregated and reflected in a single standard normal statistic,

$$\tau = \frac{1}{K}\sum_{k}\frac{\beta_{1,k}}{\sigma_{k_C}} \qquad k = 1, ..., K \tag{3.5.2}$$

where $\beta_{1,k}$ indicates the average treatment effect for outcome k and σ_{k_C} denotes the standard deviation of the k^{th} control outcome.Having included the covariates, the K average treatment effects (β_1) and sample variances can be easily acquired through a linear regression. By doing so, the above equation can be thought of as a point estimator representing a collection of standardized treatment effects. However, this paper also take account of the covariance of effects and therefore adapt a seemingly uncorrelated regression (O'Brien, 1984; Kling et al., 2007):

$$Y = I_K \otimes (T \quad X)\beta + v \tag{3.5.3}$$

where T is the treatment indicator(s), and X consists of controlled regressors as well as a constant term.

the sign of the summary index reveals information on the direction of the aggregate impact of a class of outcomes, and the more the summary index deviates from zero, the stronger is the implied aggregate effect.

Estimated effects on summary indices of mothers based on the 2012 data (panel A) and on the pooled 2012-13 data (panel B) are presented in Table 3.4. We report estimates from three different specifications, with only the treatment indictor as a covariate (column 1), with child and household demographic characteristics as controls (column 2) and with the addition of household economic characteristics as additional controls. Based on the fully specified regression (column 3), we conclude that the program enhances mother's participation in the labor market; the treatment effect on the corresponding summary index based on the 2012 data is 0.482 (se=0.098) and based on the pooled 2012-13 data it is 0.503 (se=0.098). These positive estimates are precisely measured and are practically unchanged after controlling for child and household demographics (column 2) and for economic covariates (column 3) as well. The entropy balancing estimates in columns 4-6 are very similar to the unweighted estimates presented in columns 1-3.

The corresponding treatment estimates on mothers' economic and social independence are also positive and large, 0.366 (se=0.076) when using the 2012 data and 0.276 (se=0.056) when using the pooled 2012-2013 data. A similar positive treatment effect is evident for the household decision-making outcomes, albeit the estimated coefficient for this summary index is smaller, 0.126 based on the 2012 data and 0.093 based on the 2012-13 pooled sample. The program encouraged an increase in parental investment in early child development. The estimated effect on the overall index capturing these mothers' investment is positive and significant, 0.207, se=0.065. This positive effect is seen in improvements in several measures of children's outcomes (test scores, drop-out rates, repetition). It is possible that the better education outcomes for children were the result of other improvements generated by the program, such as the increase in family income, and the direct training in cognitive and non-cognitive skills that the children received in the biweekly meetings with the family adviser. We will discuss this further when we present evidence of the effect of the *PelCa* program on children.

We note again that the estimates for these summary indices are robust to the inclusion of control variables and also to a re-weighting with entropy balancing. Taken together, these results suggest that we can gain further insights by examining in detail the effect on the mothers' outcomes. However, we do not find a strong average treatment effect on mothers' self-esteem and on personality traits (openness to experience, conscientiousness, extraversion, agreeableness, neuroticism): the estimated effect on the summary index of these aspects is small (0.045) and not precisely measured.[23] This contrasts with Kabeer (2001), who finds that access to credit improves women's self-esteem. However, our findings do not necessarily imply the absence of a treatment effect on self-esteem or personality traits, as it could also be that the instruments used to measure these outcomes were not the most appropriate.

Heterogeneous Treatment Effects

Next, to gain more insight into the treatment effects on mothers, we explore treatment heterogeneous effects. Table 3.5 presents the aggregate-estimated effect on the summary indices with control for all covariates and for subsamples of mothers based on the 2012 data. We first examine the subsamples of mothers by pre-treatment working conditions: this evidence is presented in columns 1-2 of Table 3.5. Among the group of treated mothers, 148 reported that they were working when their interviewed child was born, and 133 reported they were not working at that time. Our estimates indicate that the program had larger effect on labor market outcomes of mothers who were not working before joining *PelCa*, 0.794 (se=0.191) versus 0.340 (se=0.126), the F-test strongly rejecting that they are equal (at 5 percent significance level).

For intra-household decisions, the same pattern emerges between the two treated subsamples. Although the F-statistic is less significant, the estimated effect for mothers who were not working before *PelCa* is greater, 0.213 versus 0.086, respectively. The program has the same effect for the two groups' summary measure of economic and social independence outcomes (0.351 and 0.484, respectively). Interestingly, the average effect on child

[23] When we evaluate the effects on a summary index for fertility choices separately we find no effects.

investment comes mainly from the effect among mothers who worked at baseline (0.373) with practically no effect among mothers who did not work before enrolling in the program. The mothers with a higher pre-program employment rate might have had a more enriched upbringing, which complemented the knowledge about child rearing acquired during the program.

In columns 3-4, we present estimates by sub-samples stratified by mother's pre-program educational attainment (up to/more than primary school education). These are almost equal samples, 144 and 136, respectively. The treatment effect on the labor market index for the higher education mothers is 0.611 (se=0.163), versus 0.462 (se=0.130) in the lower education sample, both estimates being statistically significant. A similar heterogeneous pattern is seen in the effect on mother's role in family decision-making, larger for the more educated mothers, 0.205 (se=0.087) versus 0.098 (se=0.076), though the difference is not statistically significant. We find a similar effect size on the summary index of all mothers' economic and social independence outcomes for both groups (0.420 vs. 0.386). The effect on child investment is strikingly different however for the two groups, 0.320 (se=0.105) for less-educated mothers and 0.070 (se=0.100), for more educated mothers, and the difference is statistically significant. One possible explanation is that mothers with lower education levels are less skilled in child rearing and hence have higher marginal benefits from their participation in the *PelCa* program.

Next, we explored heterogeneous treatment effects by mothers' pre-treatment role in decision-making. We expect that mothers who initially (before 2005) were less involved in family decision-making would benefit more from the program.[24] The estimates in Table 3.5, columns 5-6, suggest that this group had larger gains in the labor market and in household decision-making. With respect to the latter effect, the difference is striking, 0.197 (se=0.073) versus -0.017 (se=0.083).

We next examine heterogeneity in the treatment effects by child gender. The results

[24] We made use of pre-treatment variables of mothers' intra-household decisions. After calculating the number of total household decisions that mothers made in 2005, we divide the full sample into two by the mean of total decisions. In the 2012 sample of mothers there are 179 and 93 mothers who made more decisions and fewer decisions respectively.

are presented in Table 3.6. Based on 2012 data, the treatment effect on labor market outcomes does not vary by the child gender, 0.534 for mothers of boys vs. 0.500 for mothers of girls. The effect on child investment is larger for boys (0.268 vs. 0.193) but the difference is not statistically significant. Interestingly, mothers of boys have also a larger and significant treatment effect on the score in the Rosenberg self-esteem scale test and the Big Five Personality Traits test (0.177 vs. -0.021). However, mothers of girls have larger treatment effects on intra-household decision-making (0.161, significant at 5 percent level, versus 0.079 and not significant) and on economic and social independence outcomes (0.471, significant at 1 percent level, versus 0.323, significant at 1 percent level).

In the rest of the paper we will study which specific mother's and child's outcomes drive the response of our aggregate measures. First we will present and discuss estimates based on the full sample, with and without the pre-treatment covariates. Second, we will check if our results hold when the sample is reweighted through entropy balancing. Finally, we will perform other robustness checks.

3.5.2 Estimated Effects on Mothers' Specific Outcomes

Labor-Market Outcomes

The estimates based on the 2012 data are presented in panel A of Table 3.7 and those based on the pooled 2012-13 data are presented in panel A of Table 3.8. Labor-market outcomes reflect the empowerment and emancipation of women. As we pointed out in the previous section, the effect on the summary index in Table 3.4 suggests an overall significant improvement in mothers' employability and family income, and the evidence in Tables 3.7-3.8 strengthens this conclusion. Treated mothers are 17.6 percent more likely to be working, 20.7 percent more likely to be working full-time and 20.4 percent more likely to be working in the formal sector. These estimated effects relative to the untreated mothers are an increase of 36 percent, 130 percent and 230 percent, respectively, indicating a large increase in mothers' employability.[25] Results based on the pooled 2012-13 data are

[25]The types of jobs that treated mothers hold are typically low-skilled jobs: mainly domestic cleaners, but also seamstresses and shopkeepers. More details on the job categories are available upon request.

very similar. All these estimated coefficients are significantly different from zero at the 1 percent significance level, and they are not affected by adding any of the control variables in the regression. Moreover, we find that treated mothers have more stable employment: 69 percent of the working mothers from the program were working in both 2012 and 2013, whereas only 49 percent of the working mothers in the control group were working in both years.

Another important result in Table 3.7 shows that the 2012 average monthly income of treated families is $44.48 higher than the families in the control group – a finding that we attribute largely to increased wages for mothers. The median wage of workers in the neighborhood is the minimum wage ($292 per month in 2012) and, therefore, the income gain from the program is large. In 2013 we collected the data on mothers' wages. Using this information we estimate that family income is up mainly because of an increase in mothers' wages: treated mothers earn $13.33 (se=4.87) more per week than control mothers, i.e. more than $57 extra per month (Table 3.8, panel A), while fathers' wages are not significantly different in the two groups. Our result of a significant impact on mothers' wages without any similar effect on the spouse is in contrast to the findings reported in Rosero and Oosterbeek (2011) who study the effect of child care centers on labor market outcomes of mothers. They report an estimated increase in the likelihood that a mother is working by 22 percent[26] and of household income by $80,[27] but unlike in the *PelCa* program the latter is not driven by a rise in mothers' earnings, but by the partners' income. Actually Rosero and Oosterbeek (2011) even find that home visits reduce the proportion of mothers working by 17 percent, while leaving mothers' income unaffected; however, we note that our study is based on a longer term follow up and that the long term effect of the home visits program might be different than the short term effect reported in Rosero and Oosterbeek (2011).

[26]The child care centres' effect on mothers' employability is stronger than the *PelCa* program effect (22 percent vs. 16.4 percent), but it goes in the same direction.

[27]Please note that household income in Rosero and Oosterbeek (2011) is measured in 2007, while in the current analysis it is measured in 2013, making the two incomes not perfectly comparable.

Economic and Social Independence

As seen from Table 3.4, the effect on mothers' independence from an economic and social perspective is very large. The evidence presented in panel B of Tables 3.7 and 3.8 shows that treated mothers in 2012 are 21.3 percent (se=0.056) more likely to manage their own money. Relative to the control group mean (44.7 percent), this is a 47.7 percent increase. Again, these effects are unchanged when we add to the regression each set of control variables and when we expand the sample to include also the 2013 data.

Another sign of the program's effect on women's empowerment is a larger proportion (8.1 percent more, se=0.036) of treated mothers who are studying at the survey date, which is about 155 percent higher than the rate of mothers in the control group. This estimate is significant at the 5 percent level and is unchanged even after adding all the control variables. When using the pooled 2012-13 data, the treatment effect on this outcome is 8.5 percent (se=0.031). A concern may be that entering the job market may lower the incentives and time for studying, and mothers who participated in the program were more likely to be working. To address this issue, we check whether mothers who quit studying in 2013 also found a job in 2013; we find little evidence of such correlation.

More evidence of the program's effect on women's empowerment is based on the program effect on mothers' role in the decision whether she can work. In the 2012 survey, we asked the mother who decides whether she can work. Treated mothers were 13.2 percent (se=0.044) more likely to report that this decision is taken by themselves or jointly with their partner. The estimated impact relative to the control group is 16.4 percentage points higher. This holds true when we pool data from the 2012 and 2013 surveys: treated mothers are 10.4 percent (se=0.031) more likely to have a role in this decision. Both estimates are statistically significant at the 1 percent level. Here as well, adding controls and using a weighted regression does not move the point estimates at all.

Household Decision-Making

Treated mothers are more likely to make decisions alone or with their partners about issues related to children's education (9.9 percent effect, significant at the 1 percent level) and on children's discipline (8.7 percent effect, significant at the 5 percent level), which are 11.6 and 10.4 percent greater, respectively, than the outcome means of control mothers. These estimated effects remain unchanged when controls are added. In the 2012 and 2013 pooled data, we estimated similar effects. However, we find no effect on the other domains (intra-household decisions on spending, on having children, on contraceptives and on what to do when children are ill). In the follow-up survey in 2013, we also collected intra-household decision outcomes on who decides on issues related to mothers' health, on purchasing important items and on whether mothers can visit friends and relatives. Effects on these intra-household decisions are non-conclusive. When stratifying the sample by length of participation in the program (Tables C.10-C.11), we find that mothers who have been in the program for a longer period of time are more likely to be involved in household decision-making.

Together with the results based on the single-treatment dummy, it seems safe to conclude that mothers participating in the *PelCa* program assume greater intra-household responsibilities and participate more fully in intra-household decision-making.

Access to credit was also found to have a positive effect on women's power in household decision-making, for decisions related to the loan (Kabeer, 2001, 2005). The effect of access to credit seems higher than what we find here. However, this might be due to the lower initial power of women who received the loan in the analysis by Kabeer (2001, 2005), where 20 percent of women in the comparison group had some sort of role in decision-making, compared to the women in our study, where even before the treatment, 70/80 percent of women had some power within the household.

Mothers' Child Investments

The program is intended to improve children's outcomes by enhancing investments in children's cognitive and non-cognitive skills. We therefore asked mothers in the 2012 survey how much time per week they spend interacting with their children in a variety of activities. Table 3.11 presents evidence on forms of child investment that we aggregated in the overall respective index. The *PelCa* program enhanced three of the four types of maternal investment that we considered. The estimates show that treated mothers are more likely to have conversations with children (4 percent effect, significant at 5 percent level) and also more likely to listen to and talk to their children about the books or other material they read (10 percent effect, significant at 5 percent). Treated mothers also invest more time in playing educational leisure games (for example dancing) with children (6.6 percent effect, se=0.025). We find no similar effects on mothers going to the library regularly with children (2 percent effect, se=0.040).

Taken together, the results suggest that participation in the *PelCa* program has increased mothers' investment in their children, in the form of increased mother-child educational interaction and recreational activities. In the next section we will report positive effects of the *PelCa* program on children's schooling outcomes and we view the above evidence on maternal investment as indication of a possible mechanism through which the program might have improved child development. These reduced form effects are similar to findings about the effect of a home visiting program in Colombia that increased varieties of play materials and play activities in the home and also improved children's cognitive, language and socio-emotional development (Attanasio et al., 2014). Using a structural model, Attanasio et al. (2015) provide evidence that directly links these parental investments in children to the children's improved cognitive outcomes.

Treatment Effect by Number of Years in the Program

We estimated the treatment effect where exposure is measured by number of years in the program. This specification imposes a linear effect of years in the program. We prefer this specification over estimating separate regressions using stratified samples by number of years in the program because of sample size considerations. The number of years that mothers in our sample participated in the program ranges from two to seven and the mean is 5.5 years. We note that when using a linear effect specification of years in the program as a treatment measure, the summary index cannot be calculated because the treatment variable is no longer binary. In Table C.8 and C.9 we report the linear effect estimate for each of the mothers' outcomes. Most of the estimates are positive and significantly different from zero. The pattern of positive or no effects in these two tables is consistent with the estimates we present in earlier tables. For example, the intensity of treatment measured as number of years in the program has no effect on mothers' role in decision making regarding food expenditure, having children and use of contraceptives (Table C.8), which is exactly in line with the evidence presented in Tables 3.8-3.9, where we use a simple indicator of program participation as a measure of treatment.

Next, we re-estimated the effect on mothers' and children's outcomes by two subsamples – whether the treated enrolled in the program before 2007 or from 2007. Enrollment in early years is positively correlated with the length of time mothers were in the program, and therefore it is an alternative approach to assess the effect of treatment intensity. The resulting estimates are reported in panels A of Tables C.10-C.11. Overall, the estimated effects on labor market outcomes, mothers' economic and social independence and mothers' care of children are larger for early participants than participants who joined later. For example, the effect on probability of making intra-household decisions among mothers who joined the program early is 0.195, while for later participants (from 2007) this effect is small and non-significant. The results based on the pooled sample of 2012 and 2013 present a similar pattern. The results presented in Tables C.12-C.13 suggest that the effect on children's overall summary indices for mothers who started the program before 2007 is positive and significant (0.138, se=0.063), whereas it is smaller and not significant for mothers who enrolled in the program later (0.109, se=0.094). However, we cannot reject the possibility that the two estimates are not statistically different from each other.

3.5.3 Effect on Children's Outcomes

The estimated effect of the program on children's outcomes are presented in Table 3.12. In the first row we present the program impact on summary index of children's outcomes and in rows 2-4 we present the estimates on the specific outcomes that make this index. The effect on the summary index is based on fewer observations than the sample size of each of the detailed outcomes because there are missing values for some of these outcomes, mostly for test scores, because not all the children were tested in both subjects. The estimated effect on the summary index of children's outcomes is positive and significant, 0.144 (se=0.058), suggesting that children are positively affected by participating in the *PelCa* program. We note again that this estimate is not sensitive at all to adding child, household demographics and household economics controls, nor to the entropy balancing.

The estimated effect on the average of the test scores in Spanish language and mathematics is 0.115 (se=0.103) when including all the controls. This implies that treated children have marginally higher school performance. This effect becomes clearer when we divide the children's sample by gender. Table 3.13 reports the heterogeneous effects on children's outcomes by gender and mothers' education. In columns 1 and 2, the estimated summary indices are based on 191 girls and 192 boys, separately. The *PelCa* program increased girls' average test score by 0.378 (se=0.144) and boys' by 0.012 (se=0.147). The effect on girls is larger and precisely measured and the difference between the two groups is statistically significant. The F-test statistic for this difference is 3.146, significant at 10% level. These results are different from Rosero and Oosterbeek (2011), who do not find significant gender differences in the (positive) treatment effects of home visits on children's cognitive outcomes. The High Scope Perry Preschool program however had stronger effects on female students Heckman et al. (2013).

We also estimated the program effect on this cognitive outcome when we stratified the child sample by the mother's educational levels. The results are presented in Table 3.13, columns 3-4. We divide the full sample of children into two groups according to whether their mothers had completed primary school; our data show that 207 mothers finished up to primary school, and 176 mothers have more than primary education. The estimates in columns 3-4 suggest that the effect of the program on this cognitive index is higher for children of less educated mothers (0.341, se=0.117). The respective estimated effect on children with more educated mothers is actually negative though imprecisely measured (-0.193, se=0.154). F-statistics for group differences in the treatment effects are reported in square brackets and show that the estimated impacts for the two groups are significantly different. The negative effect on children of more educated mothers could be related to the higher employment rate of these mothers before and after they joined the program and also to the lower increase in child investment during the program. We have noted above the lower gain in early childhood investment in children of educated mothers that could be explained by the small impact the program had on time inputs of these mothers into child rearing. The estimates presented in Table 3.13 columns 5-6 provide additional support for this explanation. In these columns we present program effect on stratified samples by mother's pre-program employment status and find that the effect of the program on test scores was much higher for the sample of children whose mother did not work at baseline, 0.302 (se=0.180) versus -0.092 (se=0.122).

The effect size on children's dropout rates and grade repetition presented in Table 3.12 is -0.192 (se=0.082) in a specification without covariates; it practically remained unchanged, -0.182 (se=0.073), when adding all the relevant controls, clearly indicating that treated children generally have improved educational attainment due to the program. The estimates presented in Table 3.13 suggest that impacts on girls' dropout rate or grade repetition (-0.256, se=0.089) are larger than on boys' (-0.087, se=0.116), and the difference is marginally significant. The negative effect on these outcomes is in line with evidence from other pre-school early childhood programs such as the Abecedarian project, the High Scope Perry Preschool program and the Chicago Child-Parent Centers, where grade retention is significantly lowered for treated students (Nores et al., 2005; Temple and Reynolds, 2007). The estimated effect on children's attitude towards schooling is small and imprecise (0.026, se=0.071). However, when the sample is stratified by gender, the estimate is positive on the attitude toward schooling among male participants. The estimated effect on this index is 0.142 (se=0.097) for boys, while the effect on girls is small (0.008) but with large standard error.

3.6 Robustness Checks

Evidence on Entropy Balancing: In this section we examine whether treatment effects are robust to reweighting the sample with entropy balancing. The results after imposing entropy balancing are reported in columns 4-6 of Tables 3.4, Tables 3.7-3.12 and appendix

Tables C.3-C.6. With respect to mothers' outcomes, we also calculate summary indices for the reweighted sample. The estimated effects on the indices of mothers' labor-market outcomes, independence and household decision-making exhibit similar patterns to those previously reported. The estimated impacts on summary indices tend to suggest a slightly larger overall effect on mothers' employment in 2012 (0.577 vs. 0.482 in Table 3.7). However, the 95 percent confidence interval with re-weighting (0.388, 0.766) overlaps with the confidence interval without re-weighting. The pooled data in Table 3.8 tell the same story. When it comes to disaggregated labor-market outcomes of mothers, the estimated parameters remain roughly the same: in 2012 treated mothers are 21.6 percent more likely to be working (the estimated coefficient in the reweighted sample is slightly higher than in the sample without reweighting, where we estimated a 17.6 percent increase), 23.4 percent more likely to be working full-time (20.7 percent without reweighting), and 22.0 percent more likely to be working in the formal sector (20.4 percent without reweighting). These estimates are not statistically different from the estimates in the original sample. The same pattern is observed regarding the estimated effects on average monthly family income in 2012, and mothers' wage in 2013.

The estimated effect on mother's independence does not change much either when we use the entropy re-weighting. Treated mothers are 20.5 percent more likely to control their own money, consistent with the 22.2 percent likelihood we found earlier. The effects of the program on other outcomes are also similar to previous results: treated mothers in the reweighted sample are approximately 9.5 percent more likely to be studying in 2012 or 2013 or both, and they are about 11.6 percent more likely to participate in making the decision on their job status. We find small effects on mothers' engagement in social voluntary activities. Again, the confidence interval at 1 percent for economic and social independence outcomes based on the reweighted sample overlaps greatly with the estimates using the original sample.[28]

The estimated effects on intra-household decision-making also exhibit the same pattern with entropy re-weighting: positive effect on mother's power in decisions on children's

[28]This is true both for 2012 data and the pooled data.

education and children's discipline (columns 4-6 in Tables 3.9-3.10) and no significant effect on other decision-making outcomes. The estimated aggregate effect on intra-household decision-making in 2012 is 0.124 (statistically significant at the 1 percent level), which is, again, almost identical to the previous result without re-weighting (0.126). In the pooled data, the estimated effects on the indices using the weighted data are smaller and less significant, which is consistent with our finding using the unweighted data, partly because the effects on the additional outcomes in the panel data are negligible.

The estimated effects on mothers' child investment are similar after we re-weighted the sample. The magnitudes of coefficients on the summary index and the detailed outcomes, albeit being statistically weaker, are still positive and remain within reasonable distance of the unweighted estimates.

We present the estimated effects on the children sample adjusted by entropy balancing in Table 3.12 (columns 4-6). We find the very same pattern that we found in the unadjusted data. As before, the estimated treatment impact on children's attitude towards schooling is still vague. The estimates on test scores and educational attainment are also consistent with those found in the original sample.

We also check whether improved women's status hinders men's status in the family. Reassuringly, we find no evidence of a change in the economic status of fathers with entropy-balancing either.

Attrition Analysis: A potential concern in our paper is attrition bias. Since mothers can enroll in *PelCa* soon after their child's first birthday and up to age 5, and then enroll in the *PelCa* school program, participation in the program can be up to twelve years. Mothers who remained in the program for longer may be different to mothers who decided to leave earlier. Therefore, the sample of treated mothers might be different to the mothers in the control group. To examine this potential threat to our identification strategy, we take advantage of a sample of mothers and children who enrolled in the *PelCa* program between 2004 and 2013 and left the program between 2005 and 2015. For this sample we have information on mothers' characteristics such as age, living status, working status and highest educational level, and age of their child who was enrolled in the program. This sample includes 311 mothers and 430 children.

We first restricted this "attrite" sample to 172 mothers and 258 children who joined the *PelCa* program during the same period that the treated mothers joined the program, namely between 2005 and 2009. We find that the two groups are very well balanced in terms of their pre-program characteristics (column 4 of Table C.7 in the appendix). This result is unchanged even if we restrict further the attrite sample to mothers and children who joined the program between 2005 and 2009 and left it after 4 years (column 5 of Table C.7).

One important difference between the two samples is that all treated children have reached primary school age by 2012 and moved to the school-*PelCa* program while 67 out of 258 attrite children (25.97%) were still in the pre-school *PelCa* program. We therefore limit further the attrite sample to mothers whose children were all in the school-*PelCa* program when they left the program. The means of the pre-program characteristics of the attrite group, the treatment group and the control group are presented in columns 1-3 of Table 3.15. T-tests on pre-program differences of mothers' pre-characteristics (mother's age, number of children in 2005, living status, working status and highest educational level) and in child's age before entering the program are well balanced at 10 percent level of significance. The results of this comparison between the attrite and treatment groups are presented in column 4 and the attrite and the control group in column 5. We think this evidence suggests that the decision to leave the program is not correlated with observed characteristics of mothers and children, which raises the likelihood that they are also uncorrelated with their unobserved characteristics.

Additional Robustness Checks: Our analysis relies on the similarity between treatment and control groups. However, we note that 10 percent of mother or child characteristics were not balanced at the 10 percent significance level. We interpreted this difference in pre-treatment observables characteristics as random, as might happened even in a randomly assigned treatment and control groups. We think that the close similarity between the control and treatment groups results from the similar self-selection of mothers into the two groups, as both have shown their desire to participate in the program. In support of this identification approach we show below that mothers who decided to participate in the program in 2012 did not participate before for reasons that do not affect the outcomes we study. For example, we found some imbalance in mothers working status before joining or showing interest in joining the program. We therefore re-do the analysis by subsamples: first we analyze the treatment effects in the group of treated and control mothers who were not working at baseline, and secondly in a sub-sample of treated and control mothers who were working at baseline (2005). The estimated treatment effects obtained from these two sub-samples are very similar and they are also very similar to the estimates we obtained from the full sample. This evidence rules out the possibility that our results are driven by non-comparable treatment and control groups in terms of employment rate at baseline.

It is still important, however, to understand why control group mothers did not enroll in the program previously. From the Figure 3.2 in the appendix, we can see that a high proportion of mothers did not enroll in the program because they did not know about it (44.55 percent); about a quarter of mothers because they lived too far from the NGO offices (25.74 percent); some had problems with their application forms, for example, losing the forms (11.88 percent), and some because of a previous affiliations with other NGOs in the area (8.91 percent). These responses exclude the possibility that mothers chose not to enroll in the program because they questioned its effectiveness. We also note here that none of these explanations are correlated with the previous working condition or with working full-time before.

As a further robustness check, we re-estimate all models by limiting the sample to the control mothers who did not enroll in the program either because they were already affiliated with another NGO, or because there was a problem with their application forms. The first group of control mothers, if anything, should be more attentive than treated mothers. The second group can be interpreted as randomly assigned to the control group. Unfortunately the sample here is very small (21 mothers): the estimated effects we obtain from this sample are qualitatively similar to those obtained from the full sample, but they are much less precisely estimated.

As a final robustness check, we also re-estimated the various models using propensity score matching (with replacement). The mean bias is reduced to 6.7 percent and the median bias to 4.0 percent. The sample consists of 58 treated mothers and 90 control mothers. We also estimate a propensity score matching without replacement: the mean bias is reduced to 5.4 percent and the median bias to 5.0 percent. The sample consists of 90 treated mothers and 90 control mothers. As indicated above, we prefer entropy balancing which allows for maintaining the sample size, even though when we reduce the sample size through propensity score matching, we obtain very similar results.

3.7 Conclusions

In this paper we analyze an innovative method to empower women and increase their early childhood investment in their children. *PelCa*, a home-preschool program, placed the mother at the center of her children's education, and provided guidance and training for achieving its goals. Relying on a 'designed' quasi-natural experiment, we are able to identify and measure the effects of this program on both mothers and children after several years of program participation.

First, we find that the intervention empowers women across different domains. It facilitates their entry into the labor market: treated mothers are much more likely to be working, more likely to be working full-time and more likely to be working in the formal sector. All of these estimates are precisely measured and are robust to the inclusion of covariates. Moreover, treated mothers become more financially independent and more likely to manage and make spending decisions about their own money; to be studying; and to decide whether they can work outside of the home. The results suggest that after joining the *PelCa* program, mothers are also willing to spend more time interacting with their children, on cognitive or social activities. The treatment further modifies the allocation of power in the house: mothers become more likely to take part in decisions about children's education and discipline. However, we find no effect on mother's role in decisions about what to do when the child is ill, about various types of expenditures, about having children, or the use of contraceptives. Treatment intensity plays a role here: the longer the mother stays in the program, the more empowered she becomes within the household. Moreover, mothers who were less empowered at baseline are the ones who gain the most in terms of advancing their role in decisions about the household and the allocation of resources.

All of the above results hold when we estimate aggregate treatment impacts, use summary indices instead of individual outcomes in order to account for multiple inference, when we use entropy balancing to adjust for differences in pre-treatment covariates, and when we use other robustness checks.

We also evaluate the program's impact on children. We firstly examine children's cognitive tests. The estimated impact on the summary index of cognitive achievement of treated children is positive but only marginally significant. But the average treatment effect disguises meaningful heterogeneity. Girls in the program appear to make large positive progress in test scores but there is no corresponding effect on boys; children whose mothers have a lower educational attainment benefit much more from the program in terms of improved cognitive tests than children of more educated mothers. Our findings also suggest that there are differential treatment effects by mothers' pre-treatment working status – children with less-educated mothers at baseline (before enrolling in the program) show more improvements in their cognitive testing.

Moreover, consistent with findings from other pre-school programs, students in the *PelCa* program are much less likely to drop out of school or to repeat a grade. These effects are marginally larger for female students, but the difference between genders is not significant. Although we do not find a treatment effect on students' attitudes toward schooling, the overall summary index aggregating all children's outcomes confirms that children tend to be positively affected by participating in the *PelCa* program. Overall, there is evidence that the home-preschool program that we study here helped mothers raise their children in a more learning conducive environment, which led to positive effects on children, as well as on empowering mothers at home and in the community.

Appendix:
Construction of Control Variables

Household demographics controls

1. Mothers' and fathers' education (separately): Indicator variables (0/1) for whether primary schooling not completed, primary schooling completed, secondary schooling not completed, secondary schooling completed, and university schooling completed.

2. Mother age and Father age.

3. Mother personal status: Indicator variables (0/1) for whether the mother was married, cohabited, or single.

4. Mother origin: An indicator (0/1) for whether the mother is from Quito.

5. Parents from same city: indicator (0/1).

6. Number of children in 2005.

Household economics controls

1. Mother working before child enrolled in *PelCa*: indicator=1, 0 otherwise.

2. Father working before child enrolled in *PelCa*: indicator=1, 0 otherwise.

3. Mother's firm size if employed: (0, 1, 3.5, 8, 15.5, 35.5, 75.5, 300.5).

4. Father's firm size if employed: (0, 1, 3.5, 8, 15.5, 35.5, 75.5, 300.5).

5. Family average income: before child enrollment in *PelCa* (0, 50, 200, 350, 450, 600, 800).

Child controls

1. Birth order (1,...,10).

2. Child age (5,...,14).

3. Number of young siblings in 2005 (0,...,3).

4. School fixed effects

Figures and Tables

Figure 3.1: Identification Strategy

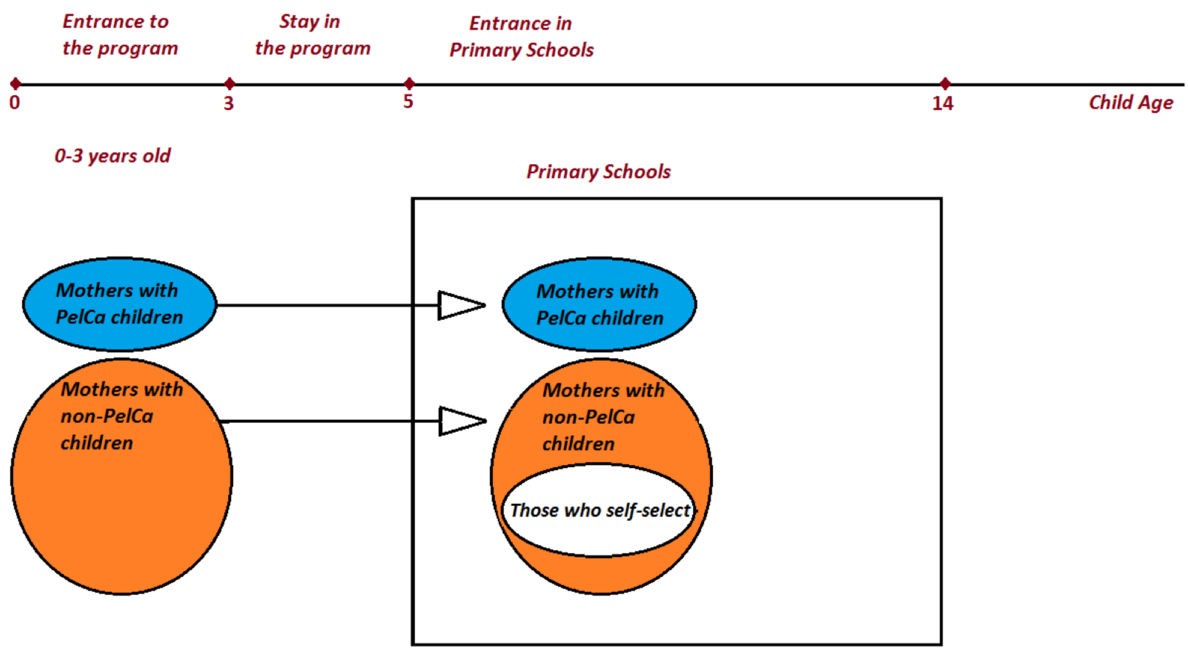


Figure 3.2: Why control mothers did not enroll before in the program

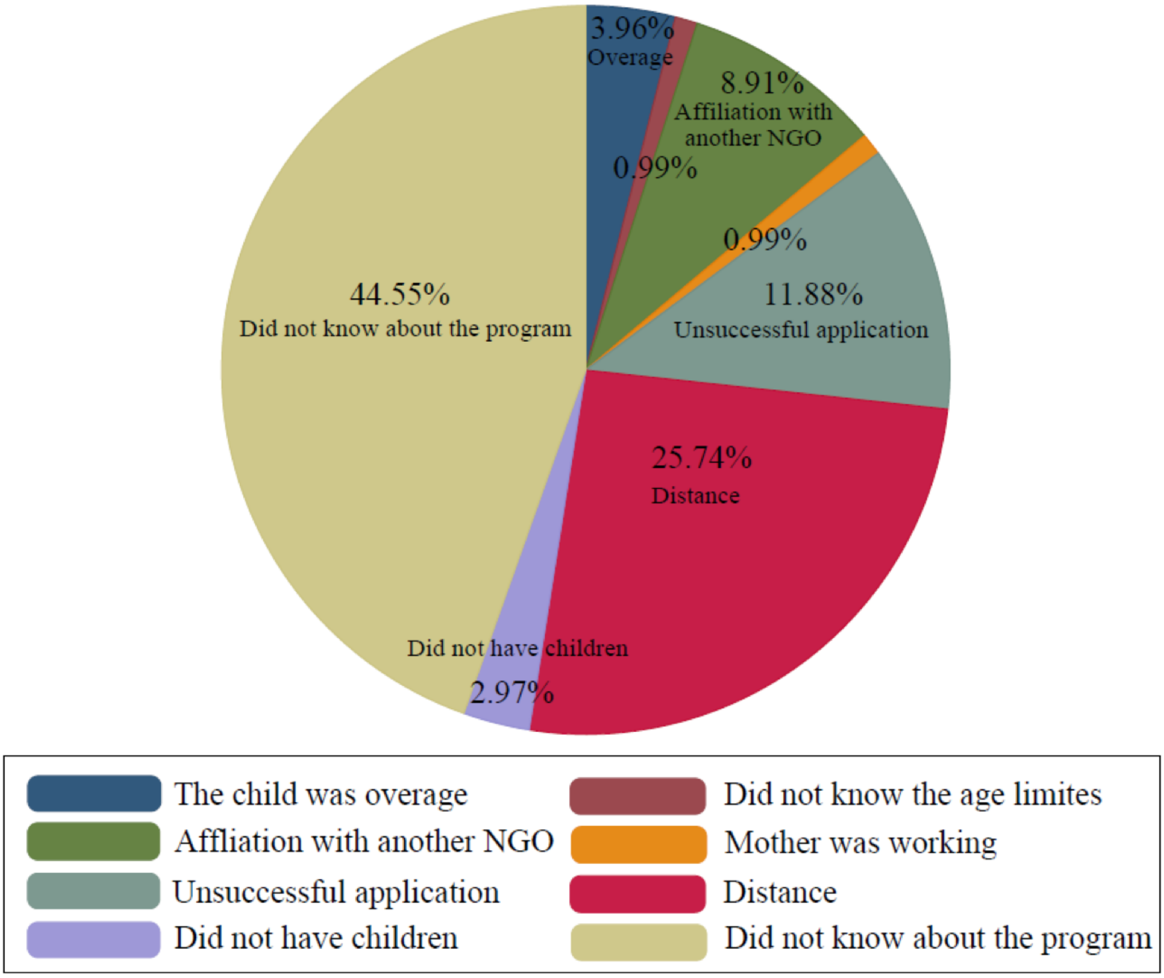


Table 3.1: MOTHERS' CHARACTERISTICS AND PRE-PROGRAM OUTCOMES (2012)

	Treatment mean	Control mean	Difference in means	Std. error
	(1)	(2)	(3)	(4)
A: Mothers' characteristics before Treatment				
Age	31.988	31.183	0.805	(0.729)
From Quito	0.560	0.496	0.065	(0.061)
Parents from same city	0.510	0.456	0.053	(0.064)
Live together with partner	0.801	0.817	-0.016	(0.048)
Divorced/separated/widow	0.018	0.009	0.009	(0.014)
Single	0.181	0.174	0.007	(0.047)
Number of children in 2005	1.849	1.632	0.218	(0.180)
Did not complete primary	0.114	0.148	-0.033	(0.041)
Completed primary	0.392	0.374	0.018	(0.059)
Did not complete secondary	0.295	0.304	-0.009	(0.056)
Completed secondary	0.169	0.165	0.003	(0.045)
Started university	0.024	0.009	0.015	(0.016)
Not religious	0.096	0.070	0.027	(0.034)
Christian	0.831	0.861	-0.030	(0.044)
B: Mothers' pre-program outcomes				
Manage own money	0.582	0.526	0.056	(0.061)
Worked	0.470	0.609	-0.139**	(0.060)
Worked full-time	0.551	0.729	-0.177**	(0.078)
Self-employed	0.808	0.786	0.022	(0.067)
Worked in the formal sector	0.256	0.300	-0.044	(0.074)
Mean firm size	10.182	12.739	-2.557	(5.855)
Reason Not Working				
Because of children	0.632	0.644	-0.012	(0.089)
Because there was no job	0.138	0.200	-0.062	(0.067)
Because partner does not want	0.138	0.133	0.005	(0.063)
Other reasons	0.092	0.022	0.070	(0.046)
Joint Decision with Spouse				
Child's education	0.899	0.875	0.024	(0.066)
Own health	0.963	0.968	-0.005	(0.040)
Discipline	0.875	0.903	-0.028	(0.069)
Expenditures	0.764	0.693	0.071	(0.054)
Food expenditures	0.758	0.789	-0.032	(0.051)
Own labor force participation	0.800	0.770	0.030	(0.050)
Having children	0.878	0.858	0.020	(0.041)
Contraceptives	0.896	0.856	0.040	(0.040)
F(23, 36) =1.2918 Prob > F = 0.2402				
Observations	166	115	281	

NOTE. The numbers attached in columns 1-3 of the last row of the table indicate the numbers of observations in the treated sample, control sample and the total sample, respectively. Statistics are based on the 2012 survey of mothers. Standard errors are presented in parentheses in the column (4); * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. An F-test on the overall significance of the pre-treatment variables is shown at the end of the table. In a later version of the paper, we have tried different specifications with control variables to ensure the robustness of the results.

Table 3.2: Children's Characteristics and Pre-program Outcomes (2012)

	Treatment mean	Control mean	Difference in means	Std. error
	(1)	(2)	(3)	(4)
Female	0.521	0.470	0.051	(0.052)
Age	8.344	8.835	-0.491**	(0.198)
Mean birth order	2.023	1.878	0.145	(0.136)
1 younger sibling in 2005	0.201	0.201	-0.000	(0.041)
2 younger siblings in 2005	0.027	0.049	-0.021	(0.019)
3 younger siblings in 2005	0.000	0.006	-0.006	(0.005)
Height at birth (cm)	48.258	48.790	-0.531	(0.462)
Weight at birth (gram)	3046.155	3029.197	16.957	(74.818)
Head circumference at birth (cm)	33.723	33.572	0.151	(0.283)
Dummy grade 1/2	0.320	0.305	0.015	(0.048)
Dummy grade 3/4	0.365	0.348	0.018	(0.050)
Dummy grade 5/6	0.242	0.268	-0.026	(0.045)
Dummy grade 7	0.073	0.079	-0.006	(0.027)
F(11, 103) = 1.3029 Prob > F = 0.2334				
Observations	219	164	383	

Note. The numbers attached in columns 1-3 of the last row of the table indicate the numbers of observations in the treated sample, control sample and the total sample, respectively. Statistics are based on the 2012 survey of children. Standard errors are presented in parentheses in the column (4); * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. An F-test on the overall significance of the pre-treatment variables is shown at the end of the table. In a later version of the paper, we have tried different specifications with control variables to ensure the robustness of the results.

Table 3.3: TREATMENT-CONTROL MEAN COMPARISONS OF OUTCOMES AND SUMMARY INDICES

	2012 sample				Pooled 2012 and 2013			
	Treatment Mean	Control Mean	Difference (1)-(2)	Std. Error	Treatment Mean	Control Mean	Difference (1)-(2)	Std. Error
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A: Labor market outcomes (mothers)								
Summary index	**0.497**	**0.000**	**0.497*** **	**(0.071)**	**0.559**	**0.000**	**0.559*** **	**(0.062)**
Works	0.693	0.487	0.206***	(0.058)	0.767	0.546	0.221***	(0.042)
Working full-time	0.378	0.157	0.222***	(0.054)	0.416	0.199	0.217***	(0.042)
Working with contract	0.279	0.087	0.192***	(0.047)	0.272	0.082	0.190***	(0.036)
Average family monthly income	344.848	300	44.848**	(20.526)	--	--	--	--
Weekly wage†	--	--	--	--	46.727	28.559	18.168***	(4.460)
B: Economic and social independence (mothers)								
Summary index	**0.319**	**0.000**	**0.319*** **	**(0.064)**	**0.251**	**0.000**	**0.251*** **	**(0.047)**
Manage own money	0.661	0.447	0.213***	(0.059)	0.681	0.487	0.194***	(0.044)
Participates in voluntary activities	0.639	0.562	0.077	(0.066)	0.584	0.550	0.034	(0.048)
Currently studying	0.133	0.052	0.081**	(0.036)	0.114	0.051	0.063**	(0.026)
Own or joint decision on own work status	0.927	0.805	0.122***	(0.039)	0.956	0.871	0.085***	(0.024)
C: Intra-household decision-making (mothers)								
Summary index	**0.093**	**0.000**	**0.093** **	**(0.043)**	**0.085**	**0.000**	**0.085*** **	**(0.028)**

cont..

Own/joint decision on child's education	0.952	0.852	0.099***	(0.034)	0.973	0.897	0.076***	(0.021)
Own/joint decision on own health	0.945	0.93	0.016	(0.029)	0.953	0.928	0.025	(0.021)
Own/joint decision on child's discipline	0.921	0.833	0.087**	(0.039)	0.939	0.856	0.083***	(0.027)
Own/joint decision on expenditure	0.805	0.754	0.05	(0.050)	0.819	0.732	0.087**	(0.038)
Own/joint decision on food expenditure	0.806	0.789	0.017	(0.049)	0.827	0.794	0.033	(0.036)
Own/joint decision on having children	0.963	0.956	0.007	(0.024)	0.959	0.958	0.001	(0.018)
Own/joint decision on contraceptives	0.920	0.946	-0.026	(0.031)	0.918	0.957	-0.039*	(0.023)
Own/joint decision on own health	--	--	--	--	0.906	0.918	-0.011	(0.026)
Own/joint decision on if mothers can visit	--	--	--	--	0.878	0.857	0.020	(0.031)
Own/joint decision on important matters	--	--	--	--	0.899	0.837	0.062**	(0.030)
D: Child's investment (mothers)								
Summary index	**0.166**	**0.000**	**0.166*** **	**(0.053)**	**--**	**--**	**--**	**--**
Talk to child (weekly)	0.988	0.965	0.022	(0.018)	--	--	--	--
Listen and talk to child about child's readings (weekly)	0.891	0.809	0.082*	(0.042)	--	--	--	--
Visit library with child (weekly)	0.115	0.087	0.028	(0.037)	--	--	--	--
Plays or dances with child (weekly)	0.988	0.928	0.060***	(0.023)	--	--	--	--
E: Self-esteem, Big Five Personality Traits and Fertility Choices (mothers)								
Summary index††	**0.030**	**0.000**	**0.030**	**(0.041)**	**--**	**--**	**--**	**--**
Rosenberg scale	3.141	3.107	0.034	(0.056)	--	--	--	--
Agreeableness	3.399	3.414	-0.015	(0.067)	--	--	--	--
Conscientiousness	3.724	3.714	0.010	(0.078)	--	--	--	--
Extraversion	3.086	3.166	-0.080	(0.066)	--	--	--	--
Neuroticism	3.044	2.964	0.080	(0.070)	--	--	--	--
Openness to Experience	3.570	3.507	0.063	(0.075)	--	--	--	--
Pregnant	0.018	0.035	-0.017	(0.019)	--	--	--	--
More children (including pregnant women)?	0.217	0.149	0.068	(0.048)	--	--	--	--
F: Labor market outcomes (fathers)								

cont..

Summary index	**-0.120**	**0.000**	**-0.120.**	**(0.076)**	**0.018**	**0.000**	**0.018.**	**(0.052)**
Working	0.953	0.970	-0.017	(0.025)	0.891	0.885	0.006	(0.030)
Working full-time	0.860	0.890	-0.03	(0.043)	0.817	0.785	0.032	(0.038)
Working with contract	0.480	0.550	-0.07	(0.065)	0.428	0.451	-0.023	(0.047)
G: Children's outcomes								
Overall summary index†††	**0.114**	**0.000**	**0.114*****	**(0.034)**	--	--	--	--
Test scores summary index	**0.117**	**0.000**	**0.117***	**(0.069)**	--	--	--	--
Language test	0.076	-0.058	0.134	(0.110)	--	--	--	--
Maths test	0.049	-0.039	0.088	(0.113)	--	--	--	--
Schooling dropout and grade repetition summary index	**-0.192**	**0.000**	**-0.192*****	**(0.056)**	--	--	--	--
Repeats at least once	0.023	0.085	-0.062***	(0.022)	--	--	--	--
Child temporarily leaves	0.014	0.049	-0.035**	(0.017)	--	--	--	--
Attitude towards schooling summary index	**0.074**	**0.000**	**0.074.**	**(0.052)**	--	--	--	--
Child likes school	0.010	-0.040	0.050	(0.104)	--	--	--	--
Child likes school (mother's perspective)	0.111	-0.237	0.348***	(0.103)	--	--	--	--
Child feels rules in the house	-0.078	0.073	-0.151	(0.104)	--	--	--	--
Child chose a book as gift	0.249	0.217	0.031	(0.045)	--	--	--	--
Observations of mothers/fathers	166	115	281		300	196	496	
Observations of children	219	164	383					

NOTE. In each panel, statistics for summary index of corresponding outcomes are reported in shading rows. Each summary index of multiple outcomes is calculated by the z-score (See section 5.1 and Kling et al., 2007 for details). Standard errors are presented in parentheses in columns (4) and (8) * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

† Estimated results for mother's weekly wage is based on the data of 2013;

†† This summary index is constructed by only outcomes of Rosenberg self-esteem scale and Big Five Personality Traits.

††† Signs of outcomes of schooling dropout and grade repetition are reversed when calculating the overall summary index of children's outcomes.

Table 3.4: Estimated Effects on Mothers' Outcomes: Summary Indices

	Not weighted			Weighted		
	(1)	(2)	(3)	(4)	(5)	(6)
A: Sample based on 2012 survey						
Labor market outcomes	0.494***	0.467***	0.482***	0.581***	0.577***	0.577***
	(0.100)	(0.099)	(0.098)	(0.114)	(0.101)	(0.096)
Economic and social independence	0.295***	0.336***	0.366***	0.249***	0.275***	0.302***
	(0.074)	(0.075)	(0.076)	(0.089)	(0.081)	(0.070)
Intra-household decision-making	0.119**	0.118**	0.126**	0.126*	0.125**	0.124**
	(0.055)	(0.056)	(0.057)	(0.072)	(0.061)	(0.055)
Child's investment	0.175***	0.209***	0.207***	0.213**	0.219**	0.212**
	(0.061)	(0.063)	(0.065)	(0.095)	(0.091)	(0.085)
Self-esteem and Big Five Personality Traits	0.038	0.033	0.045	0.060	0.060	0.063
	(0.065)	(0.067)	(0.068)	(0.083)	(0.090)	(0.083)
Observations	281	281	281	281	281	281
B: Pooled sample of 2012 and 2013						
Labor market outcomes	0.554***	0.502***	0.503***	0.660***	0.664***	0.665***
	(0.099)	(0.103)	(0.098)	(0.116)	(0.105)	(0.102)
Economic and social independence	0.231***	0.259***	0.276***	0.263***	0.285***	0.288***
	(0.054)	(0.055)	(0.056)	(0.072)	(0.061)	(0.062)
Intra-household Decision-making	0.106**	0.101**	0.093*	0.076	0.068	0.070
	(0.053)	(0.051)	(0.051)	(0.068)	(0.056)	(0.052)
Observations	496	496	496	496	496	496
Child Controls	No	Yes	Yes	No	Yes	Yes
Household Demographics	No	Yes	Yes	No	Yes	Yes
Household Economics	No	No	Yes	No	No	Yes

Note. Each cell reports the estimated mean effect from a separate regression. Column (1)-(3) present results using the original sample without entropy balancing. Column (4)-(6) stem from the weighted sample adjusted by entropy balancing. Standard errors are presented in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.5: Estimated Effects on Mothers' Outcomes by Pre-Treatment Outcomes Heterogeneity (2012)

	Mothers' pre-treatment employment status		Mothers' education		Mothers' pre-treatment role in decision-making	
	Not worked before	Worked before	Up to primary schooling	More than primary schooling	Below average	Above average
	(1)	(2)	(3)	(4)	(5)	(6)
Labor market outcomes index	0.794***	0.340***	0.462***	0.611***	0.729***	0.394***
	(0.191)	(0.126)	(0.130)	(0.163)	(0.145)	(0.131)
	[4.120]**		[0.514]		[2.650]*	
Economic and social independence index	0.351***	0.484***	0.420***	0.386***	0.482***	0.415***
	(0.120)	(0.095)	(0.105)	(0.113)	(0.118)	(0.108)
	[0.776]		[0.049]		[0.164]	
Household decisions-making index	0.213**	0.086	0.098	0.205**	0.197***	–0.017
	(0.083)	(0.079)	(0.076)	(0.087)	(0.073)	(0.083)
	[1.191]		[0.861]		[2.987]*	
Child's investment index	0.059	0.373***	0.320***	0.070	0.219***	0.186**
	(0.096)	(0.092)	(0.105)	(0.100)	(0.103)	(0.086)
	[6.755]***		[4.396]**		[0.003]	
Observations	133	148	144	136	93	179

Note. Each cell reports the estimated aggregate effect on a summary index from a separate regression based on the 2012 survey. Covariates of child characteristics, household demographics and household economics are included in regressions. In columns (1) – (6), F-test (Chow-test) statistic for subgroup difference in treatment effects are presented in square brackets. Estimated results are based on the original sample without entropy balancing. Standard errors are presented in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.6: ESTIMATED EFFECTS ON MOTHERS' OUTCOMES BY GENDER OF CHILD IN THE PROGRAM: SUMMARY INDICES (2012)

	Boy			Girl		
	(1)	(2)	(3)	(4)	(5)	(6)
Labor market outcomes	0.582***	0.493***	0.534***	0.466***	0.479***	0.500***
	(0.147)	(0.137)	(0.135)	(0.125)	(0.123)	(0.122)
Economic and social independence	0.272***	0.308***	0.323***	0.379***	0.447***	0.471***
	(0.096)	(0.094)	(0.094)	(0.107)	(0.102)	(0.105)
Intra-household decision-making	0.059	0.084	0.079	0.147**	0.128*	0.161**
	(0.078)	(0.080)	(0.082)	(0.067)	(0.068)	(0.066)
Child's Investment	0.226***	0.265***	0.268***	0.141*	0.176**	0.193**
	(0.085)	(0.087)	(0.090)	(0.076)	(0.081)	(0.081)
Self-esteem and Big Five Personality Traits	0.120	0.172*	0.177**	-0.026	-0.052	-0.021
	(0.087)	(0.088)	(0.089)	(0.083)	(0.085)	(0.083)
Child Controls	No	Yes	Yes	No	Yes	Yes
Household Demographics	No	Yes	Yes	No	Yes	Yes
Household Economics	No	No	Yes	No	No	Yes
Observations	159	159	159	165	165	165

NOTE. Each cell reports the estimated effect obtained in a separate regression without entropy balancing. Columns 1-3 present results using the sample of mother with a boy in the program and columns 4-6 present estimates for the sample of mothers to girls in the program; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.7: ESTIMATED EFFECTS ON LABOR MARKET OUTCOMES AND MOTHERS' ECONOMIC AND SOCIAL INDEPENDENCE (2012)

	Not weighted			Weighted		
	(1)	(2)	(3)	(4)	(5)	(6)
A: Labor market outcomes						
Summary index	0.494***	0.467***	0.482***	0.581***	0.577***	0.577***
	(0.100)	(0.099)	(0.098)	(0.114)	(0.101)	(0.096)
Works	0.206***	0.172***	0.176***	0.213***	0.216***	0.216***
	(0.058)	(0.061)	(0.062)	(0.081)	(0.069)	(0.066)
Working full-time	0.222***	0.205***	0.207***	0.233***	0.238***	0.234***
	(0.054)	(0.057)	(0.059)	(0.061)	(0.052)	(0.051)
Working with contract	0.192***	0.202***	0.204***	0.221***	0.223***	0.220***
	(0.047)	(0.048)	(0.049)	(0.044)	(0.041)	(0.041)
Average family monthly income	44.848**	44.538**	44.479**	44.507*	43.193*	43.242**
	(20.526)	(21.189)	(20.462)	(24.849)	(22.267)	(20.884)
B: Mothers' economic and social independence						
Summary index	0.295***	0.336***	0.366***	0.249***	0.275***	0.302***
	(0.074)	(0.075)	(0.076)	(0.089)	(0.081)	(0.070)
Manage own money	0.213***	0.233***	0.222***	0.202**	0.204***	0.205***
	(0.059)	(0.062)	(0.064)	(0.082)	(0.067)	(0.065)
Participates in voluntary activities	0.077	0.105	0.142*	0.069	0.087	0.105
	(0.066)	(0.071)	(0.072)	(0.092)	(0.076)	(0.074)
Currently studying	0.081**	0.111***	0.123***	0.087**	0.086**	0.095***
	(0.036)	(0.039)	(0.040)	(0.037)	(0.036)	(0.034)
Own or joint decision on own work status	0.122***	0.115***	0.132***	0.112*	0.108**	0.116**
	(0.039)	(0.042)	(0.044)	(0.061)	(0.054)	(0.052)
Child Controls	No	Yes	Yes	No	Yes	Yes
Household Demographics	No	Yes	Yes	No	Yes	Yes
Household Economics	No	No	Yes	No	No	Yes
Observations	281	281	281	281	281	281

NOTE. Each cell reports the estimated treatment effect from a separate regression based on the 2012 survey. Estimated summary indices of corresponding outcomes are reported in shading rows. Column 1-3 present results using the original sample without entropy balancing. Column 4-6 stem from the weighted sample adjusted by entropy balancing. Standard errors are presented in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.8: Estimated Effects on Labor Market Outcomes and Mothers' Economic and Social Independence (Pooled 2012 and 2013)

	Not weighted			Weighted		
	(1)	(2)	(3)	(4)	(5)	(6)
A: Labor market outcomes						
Summary index	0.554***	0.502***	0.503***	0.660***	0.664***	0.665***
	(0.099)	(0.103)	(0.098)	(0.116)	(0.105)	(0.102)
Works	0.221***	0.167***	0.164***	0.204**	0.204***	0.204***
	(0.056)	(0.061)	(0.060)	(0.081)	(0.068)	(0.067)
Working full-time	0.217***	0.192***	0.194***	0.240***	0.241***	0.241***
	(0.049)	(0.053)	(0.052)	(0.055)	(0.050)	(0.050)
Working with contract	0.190***	0.193***	0.193***	0.218***	0.218***	0.218***
	(0.038)	(0.038)	(0.037)	(0. 036)	(0.035)	(0.034)
Weekly wage†	18.168***	14.848***	13.331***	18.504***	17.886***	17.081***
	(4.460)	(4.852)	(4.874)	(5.713)	(4.999)	(4.966)
B: Mothers' economic and social independence						
Summary indices	0.231***	0.259***	0.276***	0.263***	0.285***	0.288***
	(0.054)	(0.055)	(0.056)	(0.072)	(0.061)	(0.062)
Manage own money	0.194***	0.187***	0.178***	0.146**	0.146***	0.146***
	(0.050)	(0.052)	(0.051)	(0.068)	(0.056)	(0.053)
Participates in voluntary activities	0.034	0.051	0.067	0.017	0.019	0.019
	(0.050)	(0.053)	(0.054)	(0.073)	(0.059)	(0.055)
Currently in school	0.063**	0.077**	0.085***	0.087***	0.088***	0.088***
	(0.028)	(0.031)	(0.031)	(0.028)	(0.028)	(0.027)
Own or joint decision on own work status	0.085***	0.093***	0.104***	0.120**	0.123***	0.125***
	(0.029)	(0.030)	(0.031)	(0.048)	(0.045)	(0.041)
Child Controls	No	Yes	Yes	No	Yes	Yes
Household Demographics	No	Yes	Yes	No	Yes	Yes
Household Economics	No	No	Yes	No	No	Yes
Observations	496	496	496	496	496	496

Note. Each cell reports the estimated treatment effect from a separate regression based on 2012 and 2013 surveys. Estimated summary indices of corresponding outcomes are reported in shading rows. Column 1-3 present results using the original sample without entropy balancing. Column 4-6 stem from the weighted sample adjusted by entropy balancing. Standard errors are presented in parentheses and are clustered at the maternal level.

† Estimated results for mothers weekly wage is based on the data of 2013.

Table 3.9: ESTIMATED EFFECTS ON MOTHERS' INTRA-HOUSEHOLD DECISION-MAKING (2012)

	Not weighted			Weighted		
	(1)	(2)	(3)	(4)	(5)	(6)
Summary index	0.119**	0.118**	0.126**	0.126*	0.125**	0.124**
	(0.055)	(0.056)	(0.057)	(0.072)	(0.061)	(0.055)
Own/joint decision on child's education	0.099***	0.099***	0.102***	0.143**	0.145**	0.143**
	(0.034)	(0.038)	(0.039)	(0.062)	(0.059)	(0.056)
Own/joint decision on own health	0.016	0.003	0.006	–0.019	–0.018	–0.018
	(0.029)	(0.032)	(0.032)	(0.025)	(0.025)	(0.026)
Own/joint decision on child's discipline	0.087**	0.081*	0.087**	0.117*	0.116*	0.115*
	(0.039)	(0.041)	(0.042)	(0.061)	(0.060)	(0.059)
Own/joint decision on expenditure	0.050	0.052	0.069	0.106	0.106	0.103
	(0.050)	(0.054)	(0.055)	(0.077)	(0.067)	(0.063)
Own/joint decision on food expenditure	0.017	-0.001	0.017	0.030	0.031	0.029
	(0.049)	(0.051)	(0.053)	(0.078)	(0.067)	(0.062)
Own/joint decision on having children	0.007	0.007	0.005	–0.004	–0.005	–0.006
	(0.024)	(0.026)	(0.027)	(0.024)	(0.023)	(0.023)
Own/joint decision on contraceptives	–0.026	–0.008	–0.019	–0.027	–0.030	–0.030
	(0.031)	(0.033)	(0.034)	(0.040)	(0.030)	(0.030)
Child Controls	No	Yes	Yes	No	Yes	Yes
Household Demographics	No	Yes	Yes	No	Yes	Yes
Household Economics	No	No	Yes	No	No	Yes
Observations	281	281	281	281	281	281

NOTE. Each cell reports the estimated treatment effect from a separate regression based on the 2012 survey. Estimated summary indices of corresponding outcomes are reported in shading rows. Column 1-3 present results using the original sample without entropy balancing. Column 4-6 estimates are based on the weighted sample adjusted by entropy balancing. Standard errors are presented in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.10: Estimated Effects on Mothers' Intra-household Decision-making (Pooled 2012 and 2013)

	Not weighted			Weighted		
	(1)	(2)	(3)	(4)	(5)	(6)
Summary index	0.106**	0.101**	0.093*	0.076	0.068	0.070
	(0.053)	(0.051)	(0.051)	(0.068)	(0.056)	(0.052)
Own/joint decision on child's education	0.076***	0.063***	0.066***	0.095**	0.095***	0.095***
	(0.025)	(0.026)	(0.026)	(0.035)	(0.031)	(0.030)
Own/joint decision on own health	0.025	0.018	0.021	0.007	0.007	0.007
	(0.032)	(0.034)	(0.033)	(0.032)	(0.031)	(0.031)
Own/joint decision on child's discipline	0.083**	0.088**	0.086**	0.119**	0.119**	0.119**
	(0.033)	(0.036)	(0.037)	(0.056)	(0.054)	(0.053)
Own/joint decision on expenditure	0.087	0.083	0.090	0.104	0.103	0.104*
	(0.055)	(0.057)	(0.057)	(0.078)	(0.064)	(0.060)
Own/joint decision on food expenditure	0.033	0.011	0.031	0.031	0.031	0.031
	(0.052)	(0.052)	(0.051)	(0.076)	(0.065)	(0.060)
Own/joint decision on important matters	0.062	0.058	0.048	0.016	0.016	0.015
	(0.045)	(0.051)	(0.052)	(0.052)	(0.047)	(0.046)
Own/joint decision on having children	0.001	0.001	–0.003	–0.018	–0.018	–0.019
	(0.026)	(0.026)	(0.027)	(0.022)	(0.021)	(0.021)
Own/joint decision on contraceptives	–0.039	–0.018	–0.024	–0.024	–0.026	–0.025
	(0.031)	(0.033)	(0.036)	(0.043)	(0.033)	(0.032)
Own/joint decision on own health	–0.011	–0.006	–0.021	–0.053*	–0.052*	–0.054*
	(0.037)	(0.040)	(0.040)	(0.031)	(0.031)	(0.030)
Own/joint decision on if mothers can visit	0.020	0.022	0.013	0.040	0.043	0.043
	(0.045)	(0.049)	(0.049)	(0.062)	(0.058)	(0.058)
Child Controls	No	Yes	Yes	No	Yes	Yes
Household Demographics	No	Yes	Yes	No	Yes	Yes
Household Economics	No	No	Yes	No	No	Yes
Observations	496	496	496	496	496	496

Note. Each cell reports the estimated treatment effect from a separate regression based on 2012 and 2013 surveys. Estimated summary indices of corresponding outcomes are reported in shading rows. Column 1-3 present results using the original sample without entropy balancing. Column 4-6 stem from the weighted sample adjusted by entropy balancing. Standard errors are presented in parentheses and are clustered at the maternal level; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.11: ESTIMATED EFFECTS ON MOTHERS' CHILD INVESTMENT (2012)

	Not weighted			Weighted		
	(1)	(2)	(3)	(4)	(5)	(6)
Summary index	0.175***	0.209***	0.207***	0.213**	0.219**	0.212**
	(0.061)	(0.063)	(0.065)	(0.095)	(0.091)	(0.085)
Talk to child (weekly)	0.022	0.038**	0.040**	0.040	0.045	0.045
	(0.018)	(0.018)	(0.019)	(0.039)	(0.034)	(0.034)
Listen and talk to child about child's readings (weekly)	0.082*	0.089*	0.100**	0.101	0.105*	0.108*
	(0.042)	(0.045)	(0.047)	(0.066)	(0.057)	(0.056)
Visit library with child (weekly)	0.028	0.025	0.020	0.054	0.048	0.047
	(0.037)	(0.039)	(0.040)	(0.040)	(0.039)	(0.038)
Plays or dances with child (weekly)	0.060***	0.068***	0.066***	0.026	0.034*	0.032*
	(0.023)	(0.024)	(0.025)	(0.018)	(0.019)	(0.018)
Child Controls	No	Yes	Yes	No	Yes	Yes
Household Demographics	No	Yes	Yes	No	Yes	Yes
Household Economics	No	No	Yes	No	No	Yes
Observations	281	281	281	281	281	281

NOTE. Each cell reports the estimated treatment effect from a separate regression based on 2012 surveys. Estimated summary indices of corresponding outcomes are reported in shading rows. Column 1-3 present results using the original sample without entropy balancing. Column 4-6 stem from the weighted sample adjusted by entropy balancing. Standard errors are presented in parentheses and are clustered at the maternal level; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.12: ESTIMATED EFFECTS ON CHILDREN'S OUTCOMES: SUMMARY INDICES (2012)

	Not weighted			Weighted		
	(1)	(2)	(3)	(4)	(5)	(6)
Overall summary index†	0.156***	0.129**	0.144**	0.124*	0.157**	0.153**
(286 observations used)	(0.056)	(0.059)	(0.058)	(0.075)	(0.073)	(0.074)
Tests scores	0.122	0.124	0.115	0.068	0.131	0.120
(322 observations used)	(0.104)	(0.099)	(0.103)	(0.155)	(0.107)	(0.106)
Schooling dropout and grade repetition	–0.192**	–0.170**	–0.182**	–0.112	–0.174*	–0.181*
(381 observations used)	(0.082)	(0.073)	(0.073)	(0.089)	(0.096)	(0.098)
Attitude towards schooling	0.080	0.027	0.026	0.106	0.072	0.054
(343 observations used)	(0.063)	(0.072)	(0.071)	(0.069)	(0.072)	(0.072)
Child Controls	No	Yes	Yes	No	Yes	Yes
Household Demographics	No	Yes	Yes	No	Yes	Yes
Household Economics	No	No	Yes	No	No	Yes
Observations	383	383	383	383	383	383

NOTE. Each cell reports the estimated effect on a summary index from a separate regression based on the 2012 survey. Column 1-3 present results using the original sample without entropy balancing. Column 4-6 stem from the weighted sample adjusted by entropy balancing. Standard errors are presented in parentheses and are clustered at the maternal level; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

† Signs of outcomes of schooling dropout and grade repetition are reversed when calculating the overall summary index.

Table 3.13: Estimated Effects on Children's Outcomes by Gender, by Parental Education and by Mothers' Employment Status (2012)

	Gender		Mothers' education		Mothers' pre-treatment employment status	
	Male	Female	Up to primary schooling	More than primary schooling	Not worked before	Worked before
	(1)	(2)	(3)	(4)	(5)	(6)
Tests scores summary index	0.012	0.378***	0.341***	–0.193	0.302**	-0.092
	(0.147)	(0.144)	(0.117)	(0.154)	(0.180)	(0.122)
	[3.146]*		[7.610]***		[3.173]**	
Schooling dropout and grade repetition summary index	–0.087	–0.256***	–0.073	-0.133*	-0.076	-0.225***
	(0.116)	(0.089)	(0.101)	(0.079)	(0.104)	(0.089)
	[1.332]		[0.219]		[1.177]	
Attitude towards schooling summary index	0.142	0.008	0.003	–0.041	-0.051	0.098
	(0.097)	(0.092)	(0.111)	(0.114)	(0.114)	(0.094)
	[0.975]		[0.077]		[0.983]	
Observations	192	191	207	176	185	198

Note. Each cell reports the estimated aggregate effect on the summary index from a separate regression based on the 2012 survey. F-test (Chow-test) statistic for subgroup difference in treatment effects is presented in square brackets. Estimated results are based on the original sample without entropy balancing. Covariates of child characteristics, household demographics and household economics are included in regressions. Standard errors are presented in parentheses and are clustered at the maternal level; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.14: Correlations between Working Condition Before Treatment and Reasons Why Control Mothers Did Not Apply

	Worked before	Working now	Worked full-time before	Working full-time now
Worked before	1			
Working now	0.2108*	1		
Did not know about the program	-0.0440	-0.0952	0.1217	-0.0616
Did not have children	-0.1049	-0.0497	0.0803	0.0838
Distance	0.0833	-0.0292	-0.1444	-0.0694
Problems with the application	-0.0306	0.1407	-0.1118	0.0921
Mother was working	0.0777	0.1051	0.0803	0.2305*
Affiliation with another NGO	0.0277	-0.0193	0.0855	-0.0405
Did not know the age limits	0.0777	-0.0952	-0.2008	-0.0434
The child was over age	-0.0519	0.0101	0.1145	0.0509

Note. Each cell reports the correlation coefficient based on the 2012 survey; * $p < 0.05$.

Table 3.15: Characteristics and Pre-program Outcomes of the Usable Sample and the Sample of Mothers/Children who Enrolled between 2005 and 2009 and All Children were in the School-PelCa when They Left

	Attrition mean	Treatment mean	Control means	Difference in means (1)-(2)	Difference in means (1)-(3)
	(1)	(2)	(3)	(4)	(5)
Mother age when enrolled	26.321	26.185	---	0.136	---
				(1.291)	
Number of children in 2005	1.974	1.849	1.632	0.125	0.342
				(0.211)	(0.237)
Mother lived together with partner	0.781	0.801	0.817	-0.020	-0.036
				(0.078)	(0.079)
Mother worked	0.493	0.470	0.609	0.023	-0.116
				(0.071)	(0.075)
Highest educational level: primary school	0.676	0.687	0.678	-0.010	-0.002
				(0.067)	(0.072)
Highest educational level: secondary school	0.250	0.169	0.165	0.081	0.085
				(0.057)	(0.061)
Highest educational level: started university	0.015	0.024	0.009	-0.009	0.006
				(0.021)	(0.016)
Family lived in Pisulli	0.579	0.675	0.583	-0.096	-0.004
				(0.066)	(0.073)
Child age when enrolled	2.903	2.696	---	0.207	---
				(0.189)	
Observations of mothers	77	162	115		
Observations of children	108	219	164		

Note. The column (1) is based on the sample of mothers/children who enrolled between 2005 and 2009 and all children were in the School-PelCa program when they left. Columns (2) and (3) are based on the 2012 survey. Standard errors are presented in parentheses in the column (4); * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Bibliography

ABADIE, A. AND G. IMBENS (2002): "Simple and bias-corrected matching estimators for average treatment effects," *NBER Working Paper No. 283*.

ABADIE, A. AND G. W. IMBENS (2008): "On the failure of the bootstrap for matching estimators," *Econometrica*, 76, 1537–1557.

——— (2012): "Bias-corrected matching estimators for average treatment effects," *Journal of Business & Economic Statistics*.

——— (2016): "Matching on the estimated propensity score," *Econometrica*, 84(2), 781–807.

ALBERT, J. H. AND S. CHIB (1993): "Bayesian analysis of binary and polychotomous response data," *Journal of the American Statistical Association*, 88(422), 669–679.

ALDOUS, D. J. (1985): "Exchangeability and related topics," in *École d'Été de Probabilités de Saint-Flour XIII—1983*, Springer, 1–198.

ALKIRE, S. (2007): "Measuring freedoms alongside wellbeing," *Well-Being in Developing Countries: New Approaches and Research Strategies*, 93–108.

ANDRABI, T., J. DAS, AND A. I. KHWAJA (2012): "What did you do all day? Maternal education and child outcomes," *Journal of Human Resources*, 47(4), 873–912.

ANGRIST, J. AND J. HAHN (2004): "When to control for covariates? Panel asymptotics for estimates of treatment effects," *Review of Economics and Statistics*, 86(1), 58–72.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2015): "Estimating the production function for human capital: Results from a randomized control trial in Colombia," *NBER Working Paper No. 20965.*

Attanasio, O. P., C. Fernández, E. O. Fitzsimons, S. M. Grantham-McGregor, C. Meghir, and M. Rubio-Codina (2014): "Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in Colombia: cluster randomized controlled trial," *BMJ*, 349, g5785.

Baker-Henningham, H. and F. López Bóo (2010): "Early childhood stimulation interventions in developing countries: A comprehensive literature review," .

Banerji, R., J. Berry, and M. Shotland (2013): "The impact of mother and participation programs on child learning: Evidence from a randomized evaluation in India," *Cambridge, MA: Abdul Latif Jameel Poverty Action Lab (J-PAL).*

Barber, S. L. and P. J. Gertler (2009): "Empowering women to obtain high quality care: Evidence from an evaluation of Mexico's conditional cash transfer programme," *Health Policy and Planning*, 24(1), 18–25.

Becker, S. O. and A. Ichino (2002): "Estimation of average treatment effects based on propensity scores," *The Stata Journal*, 2(4), 358–377.

Bernhardt, D., C. Wan, and Z. Xiao (2016): "The reluctant analyst," *Journal of Accounting Research.*

Blackwell, D. and J. B. MacQueen (1973): "Ferguson distributions via Pólya urn schemes," *Annals of Statistics*, 353–355.

Brier, G. W. (1950): "Verification of forecasts expressed in terms of probability," *Monthly Weather Review*, 78(1), 1–3.

Campbell, F. A., C. T. Ramey, E. Pungello, J. Sparling, and S. Miller-Johnson (2002): "Early childhood education: Young adult outcomes from the Abecedarian Project," *Applied Developmental Science*, 6(1), 42–57.

CARNEIRO, P., C. MEGHIR, AND M. PAREY (2013): "Maternal education, home environments, and the development of children and adolescents," *Journal of the European Economic Association*, 11(s1), 123–160.

DEHEJIA, R. H. AND S. WAHBA (1999): "Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs," *Journal of the American Statistical Association*, 94, 1053–1062.

——— (2002): "Propensity score-matching methods for nonexperimental causal studies," *Review of Economics and Statistics*, 84(1), 151–161.

DIAMOND, A. AND J. S. SEKHON (2012): "Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies," *Review of Economics and Statistics*.

DUFLO, E. (2012): "Women empowerment and economic development," *Journal of Economic Literature*, 50(4), 1051–1079.

DUNSON, D. (2008): "Lecture notes in generalized linear models," .

FERGUSON, T. S. (1973): "A Bayesian analysis of some nonparametric problems," *Annals of Statistics*, 209–230.

——— (1983): "Bayesian density estimation by mixtures of normal distributions," *Recent Advances in Statistics*, 24, 287–302.

FIRTH, D. (1993): "Bias reduction of maximum likelihood estimates," *Biometrika*, 80(1), 27–38.

GELMAN, A., A. JAKULIN, M. G. PITTAU, AND Y.-S. SU (2008): "A weakly informative default prior distribution for logistic and other regression models," *Annals of Applied Statistics*, 1360–1383.

GERTLER, P., J. HECKMAN, R. PINTO, A. ZANOLINI, C. VERMEERSCH, S. WALKER, S. M. CHANG, AND S. GRANTHAM-MCGREGOR (2014): "Labor market returns to an early childhood stimulation intervention in Jamaica," *Science*, 344(6187), 998–1001.

HAHN, J. (1998): "On the role of the propensity score in efficient semiparametric estimation of average treatment effects," *Econometrica*, 315–331.

HAINMUELLER, J. (2011): "Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies," *Political Analysis*, 25–46.

HECKMAN, J., S. H. MOON, R. PINTO, P. SAVELYEV, AND A. YAVITZ (2010): "Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program," *Quantitative Economics*, 1(1), 1–46.

HECKMAN, J., R. PINTO, AND P. SAVELYEV (2013): "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes," *American economic review*, 103(6), 2052–2086.

HECKMAN, J. AND R. ROBB (1985): "Using longitudinal data to estimate age, period and cohort effects in earnings equations," in *Cohort Analysis in Social Research*, Springer, 137–150.

HECKMAN, J. J., H. ICHIMURA, AND P. E. TODD (1997): "Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme," *Review of Economic Studies*, 64, 605–654.

HEINZE, G. (2006): "A comparative investigation of methods for logistic regression with separated or nearly separated data," *Statistics in Medicine*, 25(24), 4216–4226.

HEINZE, G. AND M. SCHEMPER (2002): "A solution to the problem of separation in logistic regression," *Statistics in Medicine*, 21(16), 2409–2419.

HOFF, P. D. (2006): "Introduction to Bayesian Statistics for the Social Sciences," *University of Washington, Washington.*

HOLE, A. R. (2015): "Ranking economics journals using data from a national research evaluation exercise," *The Sheffield Economic Research Paper Series (SERPS) No. 201501.*

IMBENS, G. W. (2000): "The role of the propensity score in estimating dose-response functions," *Biometrika*, 87(3), 706–710.

JARA, A., M. J. GARCÍA-ZATTERA, AND E. LESAFFRE (2007): "A Dirichlet process mixture model for the analysis of correlated binary responses," *Computational Statistics & Data Analysis*, 51(11), 5402–5415.

JUMP, P. (2014): "REF 2014 results: Table of excellence," *Times Higher Education.*

KABEER, N. (2001): "Conflicts over credit: Re-evaluating the empowerment potential of loans to women in rural Bangladesh," *World Development*, 29(1), 63–84.

——— (2005): "Gender equality and women's empowerment: A critical analysis of the third millennium development goal 1," *Gender & Development*, 13(1), 13–24.

KAGITCIBASI, C., D. SUNAR, S. BEKMAN, N. BAYDAR, AND Z. CEMALCILAR (2009): "Continuing effects of early enrichment in adult life: The Turkish Early Enrichment Project 22 years later," *Journal of Applied Developmental Psychology*, 30(6), 764–779.

KIERNAN, K. E. AND M. C. HUERTA (2008): "Economic deprivation, maternal depression, parenting and children's cognitive and emotional development in early childhood," *British Journal of Sociology*, 59(4), 783–806.

KIM, S., N. SHEPHARD, AND S. CHIB (1998): "Stochastic volatility: likelihood inference and comparison with ARCH models," *Review of Economic Studies*, 65(3), 361–393.

KLING, J. R., J. B. LIEBMAN, AND L. F. KATZ (2007): "Experimental analysis of neighborhood effects," *Econometrica*, 75(1), 83–119.

KLUGMAN, J., L. HANMER, S. TWIGG, T. HASAN, J. MCCLEARY-SILLS, AND J. SANTAMARIA (2014): *Voice and Agency: Empowering Women and Girls for Shared Prosperity*, Washington, DC: World Bank.

KOOP, G., D. J. POIRIER, AND J. L. TOBIAS (2007): *Bayesian Econometric Methods*, Cambridge University Press.

LaLonde, R. J. (1986): "Evaluating the econometric evaluations of training programs with experimental data," *American Economic Review*, 604–620.

LeVine, R., S. LeVine, and B. Schnell (2001): ""Improve the women": Mass schooling, female literacy, and worldwide social change," *Harvard Educational Review*, 71(1), 1–51.

Li, M. (2013): "Using the Propensity Score Method to Estimate Causal Effects A Review and Practical Guide," *Organizational Research Methods*, 16(2), 188–226.

Liu, J. (2001): "Monte Carlo strategies in statistical computing," Springer, New York.

Mocan, N. H. and C. Cannonier (2012): "Empowering women through education: Evidence from Sierra Leone," *NBER Working Paper No. 18016.*

Narayan-Parker, D. (2005): *Measuring Empowerment: Cross-Disciplinary Perspectives*, World Bank Publications.

Neal, R. M. (2000): "Markov chain sampling methods for Dirichlet process mixture models," *Journal of Computational and Graphical Statistics*, 9(2), 249–265.

Nores, M. and W. S. Barnett (2010): "Benefits of early childhood interventions across the world:(Under) Investing in the very young," *Economics of Education Review*, 29(2), 271–282.

Nores, M., C. R. Belfield, W. S. Barnett, and L. Schweinhart (2005): "Updating the economic impacts of the High/Scope Perry Preschool program," *Educational Evaluation and Policy Analysis*, 27(3), 245–261.

Norets, A. and J. Pelenis (2012): "Bayesian modeling of joint and conditional distributions," *Journal of Econometrics*, 168(2), 332–346.

O'Brien, P. C. (1984): "Procedures for comparing samples with multiple endpoints," *Biometrics*, 1079–1087.

Oyitso, M. and C. Olomukoro (2012): "Enhancing women's development through literacy education in Nigeria," *Review of European Studies*, 4(4), 66.

Pidd, M. and J. Broadbent (2015): "Business and management studies in the 2014 Research Excellence Framework," *British Journal of Management*, 26(4), 569–581.

Pitman, J. (2002): "Combinatorial stochastic processes," Notes for Saint Flour Summer School.

Pitt, M., D. Chan, and R. Kohn (2006): "Efficient Bayesian inference for Gaussian copula regression models," *Biometrika*, 93(3), 537–554.

Rasmussen, C. E. and Z. Ghahramani (2002): "Infinite mixtures of Gaussian process experts," *Advances in Neural Information Processing Systems*, 2, 881–888.

Rosenbaum, P. R. and D. B. Rubin (1983): "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70(1), 41–55.

——— (1984): "Reducing bias in observational studies using subclassification on the propensity score," *Journal of the American Statistical Association*, 79(387), 516–524.

Rosero, J. and H. Oosterbeek (2011): "Trade-offs between different early childhood interventions: Evidence from Ecuador," *Tinbergen Institute Discussion Paper*.

Rossi, P. (2013): "Bayesian semi-parametric and non-parametric methods with applications to marketing and micro-econometrics," .

Rubin, D. B. and E. A. Stuart (2006): "Affinely invariant matching methods with discriminant mixtures of proportional ellipsoidally symmetric distributions," *Annals of Statistics*, 1814–1826.

Rubin, D. B. and N. Thomas (1992a): "Affinely invariant matching methods with ellipsoidal distributions," *Annals of Statistics*, 1079–1093.

——— (1992b): "Characterizing the effect of matching using linear propensity score methods with normal distributions," *Biometrika*, 79(4), 797–809.

——— (1996): "Matching using estimated propensity scores: Relating theory to practice," *Biometrics*, 249–264.

SCImago (2007): "SJR — SCImago Journal and Country Rank," *http://www.scimagojr.com/*, Retrieved July 21, 2015.

Sen, P. (1999): "Enhancing women's choices in responding to domestic violence in Calcutta: A comparison of employment and education," *European Journal of Development Research*, 11(2), 65–86.

Smith, J. A. and P. E. Todd (2005): "Does matching overcome LaLonde's critique of nonexperimental estimators?" *Journal of econometrics*, 125(1), 305–353.

Smith, M. S. and M. A. Khaled (2012): "Estimation of copula models with discrete margins via Bayesian data augmentation," *Journal of the American Statistical Association*, 107(497), 290–303.

Stock, J. H. and M. W. Watson (2007): "Why has US inflation become harder to forecast?" *Journal of Money, Credit and Banking*, 39(s1), 3–33.

Stuart, E. A. (2010): "Matching methods for causal inference: A review and a look forward," *Statistical Science: A Review Journal of the Institute of Mathematical Statistics*, 25(1), 1.

Teh, Y. W. (2011): "Dirichlet process," in *Encyclopedia of Machine Learning*, Springer, 280–287.

Teh, Y. W., M. I. Jordan, M. J. Beal, and D. M. Blei (2012): "Hierarchical dirichlet processes," *Journal of the American Statistical Association.*

Temple, J. A. and A. J. Reynolds (2007): "Benefits and costs of investments in preschool education: Evidence from the Child–Parent Centers and related programs," *Economics of Education Review*, 26(1), 126–144.

WATANABE, K., R. FLORES, J. FUJIWARA, AND L. T. H. TRAN (2005): "Early childhood development interventions and cognitive development of young children in rural Vietnam," *Journal of Nutrition*, 135(8), 1918–1925.

WONG, Y. N. (2012): "World Development Report 2012: Gender equality and development," in *Forum for Development Studies*, Taylor & Francis, vol. 39(3), 435–444.

WOOD, F. AND M. J. BLACK (2008): "A nonparametric Bayesian alternative to spike sorting," *Journal of Neuroscience Methods*, 173(1), 1–12.

ZORN, C. (2005): "A solution to separation in binary response models," *Political Analysis*, 13(2), 157–170.

A Appendices to Chapter 1

The appendix to the paper "*How does the REF panel perceive journals? A new approach to estimating ordinal response model with censored outcome*" consists of three parts. Appendix A.1 provides details of the calculation of the posterior means and posterior variances of the GPA scores. In Appendix A.2, we provide detailed descriptive statistics based on information from the REF 2014 assessment results and the SCImago Journal & Country Rank in 2014. In Appendix A.3, we provide various plots for summarising and investigating the MCMC posterior distributions.

A.1 Predicted GPA Scores of Each Institution

Let the individual output be labelled as i and the institution as j. Y_i denotes the outcome variable. We are interested in the GPA score for each institution, denoted as GPA_j. This can be written as the average GPA score of all individual outputs from the same institution:

$$GPA_j = \frac{1}{n_j} \sum_{i=1}^{n_j} GPA_{ij}$$

where GPA_{ij} indicates a particular output i from institution j and can be expressed as a weighted average quantity by its definition:

$$GPA_{ij} = 1 \times I(Y_i = 1) + 2 \times I(Y_i = 2) + 3 \times I(Y_i = 3) + 4 \times I(Y_i = 4)$$

Before calculating the mean and variance of GPA_{ij}, we note that each indicator function $\{I(y_i = s), s = 1, ..., 4\}$ in the above equation is essentially a Bernoulli variable. A vector of the indicator functions hence forms a Multinomial distribution. In this spirit, we assume that the random vector $Z_i = \Big(I(Y_i = 1), I(Y_i = 2), I(Y_i = 3), I(Y_i = 4)\Big)'$ follows a Multinomial distribution with a probability vector $\big(p_{Y_i,1}, p_{Y_i,2}, p_{Y_i,3}, p_{Y_i,4}\big)'$, where $\{p_{Y_i,s}, s = 1, 2, 3, 4\}$ indicates the probability of Y_i being equal to s. Since Z_i corresponds to a single output, the number of the trial parameter of the Multinomial distribution is exactly one. Thus, $Z_i \sim \mathbf{Mult}\big(1, (p_{Y_i,1}, p_{Y_i,2}, p_{Y_i,3}, p_{Y_i,4})'\big)$ has the following mean and (co)variance expressions:

$$\begin{aligned}
\mathrm{E}(Z_i) &= \big[p_{Y_i,1}, p_{Y_i,2}, p_{Y_i,3}, p_{Y_i,4}\big]' \\
\mathrm{Var}(Z_i) &= \big[p_{Y_i,1}(1 - p_{Y_i,1}), p_{Y_i,2}(1 - p_{Y_i,2}), p_{Y_i,3}(1 - p_{Y_i,3}), p_{Y_i,4}(1 - p_{Y_i,4})\big]' \\
\mathrm{Cov}_{q,r} &= -p_{Y_i,q}p_{Y_i,r} \qquad \text{for the } q^{th} \text{ and } r^{th} \text{ marginals of } Z_i\text{; and } q \neq j
\end{aligned}$$

By using the facts above, the mean and the variance of the GPA_j for institution j can be written as:

$$\begin{aligned}
\mathrm{E}(GPA_j) &= \frac{1}{n_j}\sum_{i=1}^{n_j}\mathrm{E}(GPA_{ij}) \\
&= \frac{1}{n_j}\sum_{i=1}^{n_j} p_{Y_i,1} + 2p_{Y_i,2} + 3p_{Y_i,3} + 4p_{Y_i,4} \\
\mathrm{Var}(GPA_j) &= \frac{1}{n_j^2}\sum_{i=1}^{n_j}\mathrm{Var}(GPA_{ij}) \\
&= \frac{1}{n_j^2}\sum_{i=1}^{n_j}\Big(p_{Y_i,1}(1 - p_{Y_i,1})\} + 4p_{Y_i,2}(1 - p_{Y_i,2})\} \\
&\quad +9p_{Y_i,3}(1 - p_{Y_i,3})\} + 16p_{Y_i,4}(1 - p_{Y_i,4})\} \\
&\quad -2\big[2p_{Y_i,1}p_{Y_i,2} + 3p_{Y_i,1}p_{Y_i,3} + 4p_{Y_i,1}p_{Y_i,4} \\
&\quad +6p_{Y_i,2}p_{Y_i,3} + 8p_{Y_i,2}p_{Y_i,4} + 12p_{Y_i,3}p_{Y_i,4}\big]\Big)
\end{aligned}$$

where $\{p_{Y_i,s}, s = 1, 2, 3, 4\}$ in the above equation are estimated as a function of the posterior of $\theta = (\beta, c)'$:

$$\begin{aligned} p_{Y_i,s} &\coloneqq \Pr(Y_i = s|\theta) \\ &= \Phi(c_s - \beta' D_i) - \Phi(c_{s-1} - \beta' D_i) \qquad \text{for} \quad s = 1, 2, 3, 4 \text{ and } \quad i = 1, ..., n_j \end{aligned}$$

A.2 Data Descriptive Statistics

Table A.1: Number of Submissions to the REF Economics and Econometrics Sub-Panel and the Business and Management Studies Sub-Panel

Journal Title	SJR Ranking in 2014	Number of Submissions to REF panels: Economics and Econometrics	Number of Submissions to REF panels: Business and Management Studies
	(1)	(2)	(3)
Quarterly Journal of Economics	6	29	6
Nature	11	2	7
Econometrica	18	69	9
Journal of Political Economy	20	22	2
Journal of Economic Literature	22	4	4
Journal of Finance	23	6	66
New England Journal of Medicine	41	1	1
Review of Economic Studies	47	63	13
Review of Financial Studies	48	5	87
Science	49	3	0
Lancet, The	55	2	5
Journal of Financial Economics	56	5	76
American Economic Journal: Applied Economics	65	16	0
American Political Science Review	75	1	2
American Economic Review	83	108	28
Journal of Economic Perspectives	109	3	4
American Economic Journal: Economic Policy	114	12	2
Annual Review of Economics	127	1	1
Annals of Statistics	131	3	0
Journal of the European Economic Association	134	71	11
American Journal of Political Science	148	1	0
Proceedings of the National Academy of Sciences of the United States	155	3	4
American Economic Journal: Macroeconomics	159	11	1
Review of Economics and Statistics	175	59	23
Theoretical Economics	204	13	1
Journal of International Economics	236	36	13
Brookings Papers on Economic Activity	240	1	0
Journal of Business and Economic Statistics	242	13	15
Journal of Labor Economics	246	13	6
Journal of Economic Growth	248	4	3
Economic Policy	252	6	5
Quantitative Economics	262	11	1
Econometric Theory	285	35	6
Journal of Econometrics	290	93	26
Journal of Monetary Economics	292	42	13
American Economic Journal: Microeconomics	293	14	4
Review of Economic Dynamics	295	14	3
Journal of Financial and Quantitative Analysis	298	3	25
Economic Journal	308	103	43
Journal of the American Statistical Association	334	5	5
Journal of Economic Theory	335	82	11
Biometrika	338	2	3
Experimental Economics	353	8	3
Journal of Public Economics	358	57	8
RAND Journal of Economics	362	16	5
Journal of Human Resources	374	7	1
International Economic Review	375	28	4
Quantitative Marketing and Economics	394	1	1
Stroke	408	1	0

cont..

Annual Review of Financial Economics	433	1	0
Journal of Development Economics	439	48	5
Management Science	450	6	56
Journal of Applied Econometrics	463	24	14
Cognitive Psychology	465	1	0
Global Environmental Change	491	1	8
Review of Finance	495	2	27
IEEE Transactions on Information Theory	509	2	0
Journal of Conflict Resolution	542	1	0
Journal of Urban Economics	545	3	7
Journal of Public Administration Research and Theory	547	1	16
Advances in Mathematics	575	1	0
Games and Economic Behavior	580	78	13
Demography	661	2	3
Proceedings of the London Mathematical Society	674	1	0
Journal of Economic Geography	689	5	35
Journal of Environmental Economics and Management	744	11	12
Journal of Financial Markets	749	1	3
Journal of Regional Science	832	1	3
Research Policy	838	1	145
Econometrics Journal	849	9	5
European Journal of Operational Research	860	1	167
Climatic Change	881	1	2
Economic Theory	883	48	14
Journal of Health Economics	917	33	24
Journal of Population Economics	955	9	9
European Economic Review	1010	51	24
Journal of Risk and Uncertainty	1039	8	16
Foundations of Computational Mathematics	1163	1	0
Journal of Industrial Economics	1191	6	7
Journal of Human Capital	1208	2	1
Journal of Money, Credit and Banking	1218	32	37
Economics and Human Biology	1256	1	1
Evolution and Human Behavior	1304	1	0
Energy Policy	1316	3	20
Energy Economics	1329	5	13
Asia Pacific Journal of Management	1347	1	4
Finance and Stochastics	1355	1	0
Statistics and Computing	1374	1	2
World Bank Economic Review	1407	5	1
Journal of Economics and Management Strategy	1466	8	6
Ecological Economics	1479	6	21
Journal of Law and Economics	1527	4	0
British Educational Research Journal	1531	2	5
Medical Decision Making	1637	1	0
International Journal of Industrial Organization	1644	16	17
Health Economics	1696	4	7
Journal of Behavioral Decision Making	1721	1	7
PLoS One	1776	1	5
Statistical Methods in Medical Research	1822	1	1
Journal of Financial Stability	1918	4	4
Journal of Combinatorial Theory - Series A	1924	2	0
Economic History Review	1952	9	25
Journal of Common Market Studies	1967	2	8
Environmental and Resource Economics	1973	15	3
Quarterly Journal of Political Science	1995	2	0
International Journal of Forecasting	2017	5	46

cont..

Econometric Reviews	2030	2	1
Electoral Studies	2070	1	0
Journal of the Royal Statistical Society. Series A: Statistics in Society	2077	19	14
Review of Economics of the Household	2116	1	0
Economics of Education Review	2151	2	11
Economic Inquiry	2161	13	18
Journal of Economic History	2177	12	5
Energy Journal	2198	4	15
Journal of Economic Dynamics and Control	2232	44	18
European Journal of Political Economy	2234	3	5
Economica	2280	18	16
World Development	2349	7	29
Journal of Financial Econometrics	2367	1	5
Public Choice	2414	14	14
Small Business Economics	2417	1	31
Journal of the Royal Statistical Society. Series C: Applied Statistics	2442	2	0
Journal of Banking and Finance	2476	23	183
Computational Statistics and Data Analysis	2546	7	4
European Review of Economic History	2632	4	1
Urban Studies	2646	1	27
Journal of Time Series Analysis	2657	9	6
Scandinavian Journal of Economics	2665	14	5
Labour Economics	2674	19	19
Area	2735	1	0
Regional Science and Urban Economics	2739	1	10
Mathematics of Operations Research	2761	1	6
American Journal of Agricultural Economics	2811	2	4
Canadian Journal of Economics	2818	24	6
Journal of Law, Economics, and Organization	2857	11	4
Journal of Empirical Finance	2878	5	23
British Journal of Industrial Relations	2917	3	102
Journal of Economic Behavior and Organization	2934	42	46
Food Policy	2997	2	2
Journal of Economic Psychology	3007	2	5
Spatial Economic Analysis	3012	2	2
Journal of International Money and Finance	3066	15	40
Regional Studies	3134	2	76
Journal of Environmental Management	3170	1	2
Oxford Bulletin of Economics and Statistics	3210	28	21
Review of International Economics	3270	6	14
Canadian Journal of Statistics	3286	1	0
Economic Development and Cultural Change	3317	2	0
Extremes	3346	1	1
Journal of Economic Surveys	3360	1	2
Applied Numerical Mathematics	3382	1	0
Journal of Symbolic Computation	3427	1	0
Journal of Computational and Applied Mathematics	3445	1	0
Journal of Nonparametric Statistics	3501	1	0
Journal of Statistical Planning and Inference	3514	5	0
Discrete Mathematics	3546	1	0
Journal of Mathematical Economics	3624	18	4
Journal of Algebraic Combinatorics	3643	1	0
International Tax and Public Finance	3818	3	4
Advances in Computational Mathematics	3850	1	0
B.E. Journal of Theoretical Economics	3871	5	0
Journal of Agricultural Economics	3890	10	5

cont..

Explorations in Economic History	3918	11	7
Journal of Public Economic Theory	4020	12	2
Social Choice and Welfare	4073	14	1
European Review of Agricultural Economics	4100	2	2
Journal of Business Finance and Accounting	4130	1	77
Oxford Review of Economic Policy	4147	2	3
SIAM Journal on Applied Mathematics	4231	1	0
Theory and Decision	4266	13	4
Ophthalmic Epidemiology	4305	1	0
Review of World Economics	4308	1	1
Physical Review E - Statistical, Nonlinear, and Soft Matter Physics	4335	1	0
Electronic Journal of Combinatorics	4387	1	0
B.E. Journal of Economic Analysis and Policy	4418	3	2
Kyklos	4500	2	3
Knowledge Engineering Review	4559	1	0
Journal of Empirical Legal Studies	4692	1	1
Journal of Forecasting	4739	3	25
Macroeconomic Dynamics	4771	10	2
Economics and Politics	4821	2	2
Advances in Mathematics of Communications	4835	1	0
Journal of Development Studies	4836	1	12
World Economy	4845	1	5
Journal of Evolutionary Economics	4905	1	4
Experimental Mathematics	4972	1	0
Few-Body Systems	5067	1	0
International Review of Finance	5078	1	2
Journal of Comparative Economics	5204	8	17
Journal of Productivity Analysis	5287	1	7
Journal of Approximation Theory	5424	1	0
Crop Science	5432	4	1
Annals of Finance	5542	5	3
Statistics and Probability Letters	5690	1	0
Journal of Group Theory	5835	1	0
Economics Letters	5892	62	76
Oxford Economic Papers	5894	24	21
International Journal of Theoretical and Applied Finance	5925	1	0
European Physical Journal B	6016	1	0
B.E. Journal of Macroeconomics	6020	9	0
Cambridge Journal of Economics	6125	4	45
Journal of Physics A: Mathematical and Theoretical	6146	1	0
Journal of Macroeconomics	6195	3	8
Agricultural Economics	6264	1	2
Resources Policy	6270	1	0
Mathematics and Financial Economics	6328	2	0
Review of Income and Wealth	6356	6	3
Mathematical Social Sciences	6407	7	3
International Review of Economics and Finance	6509	1	3
Journal of Mathematical Cryptology	6510	1	0
North American Journal of Economics and Finance	6631	2	0
International Journal of Game Theory	6681	10	1
Journal of Policy Modeling	6689	2	3
Calcolo	6978	1	0
International Journal of Economic Theory	7014	1	0
Geneva Papers on Risk and Insurance: Issues and Practice	7130	1	0
Southern Economic Journal	7244	5	7

cont..

CESifo Economic Studies	7326	1	0
Review of Economic Design	7520	1	0
Applied Economics	7795	6	11
Journal of Economic Studies	7909	1	1
International Journal of the Economics of Business	7966	1	14
Open Economies Review	8056	1	1
Manchester School	8203	10	26
Journal of Sports Economics	8342	1	0
Artificial Life	8436	1	0
Economic Modelling	8455	7	11
International Review of Law and Economics	8489	1	3
Empirical Economics	8556	1	5
Journal of International Financial Markets, Institutions and Money	8557	6	50
Journal of Post Keynesian Economics	8594	1	0
Economics and Philosophy	8621	1	0
Journal of Economics/ Zeitschrift fur Nationalokonomie	8939	1	1
International Journal of Finance and Economics	8979	1	2
Journal of Cultural Economics	9158	1	1
Economic Record	9188	2	3
Studies in Nonlinear Dynamics and Econometrics	9336	6	5
Scottish Journal of Political Economy	9375	2	8
European Journal of Finance	9378	2	97
Asian Economic Papers	9396	1	0
Computational Statistics	9468	2	1
Review of Quantitative Finance and Accounting	9488	2	25
Journal of International Trade and Economic Development	9528	1	0
Education Economics	9629	1	0
Economics of Governance	9748	1	1
International Journal of Accounting	10020	1	11
Journal of Socio-Economics	10680	1	3
Fiscal Studies	10709	1	0
International Review of Financial Analysis	10795	1	76
Review of Law and Economics	11456	1	0
Business History	11564	1	97
Journal of Institutional and Theoretical Economics	11700	3	2
Emerging Markets Finance and Trade	11937	1	0
Journal of Chinese Economic and Business Studies	12112	2	1
Journal of Pension Economics and Finance	12228	2	0
Applied Economics Letters	12259	3	1
Economics	12383	2	1
Applied Financial Economics	12400	1	20
Review of Financial Economics	13262	1	0
Hitotsubashi Journal of Economics	14652	1	0
Economics Bulletin	15272	1	1
International Game Theory Review	15960	1	0
Bulletin of Economic Research	16142	3	8
Australian Economic Papers	16866	1	1
Economie et Statistique	18421	1	0
Business & management: other submissions		---	9115
Economics & Econometrics: unpublished		182	---
Economics & Econometrics: book or book charpters		32	---
Economics & Econometrics: other articles		15	---
Total submissions		2600	12170

Source.— Column (1) uses data from the SCImago Journal & Country Rank published in 2014. Column (2) uses data from the REF 2014 Economics and Econometrics Assessment Unit, and column (3) the REF 2014 Business and Management Studies Assessment Unit.

Table A.2: List of Abbreviations of Economics Journals in this Paper

AEJ	American Economic Journal
AER	American Economic Review
EER	European Economic Review
EJ	Economic Journal
IER	International Economic Review
JAE	Journal of Applied Econometrics
JDE	Journal of Development Economics
JEEA	Journal of European Economic Association
JET	Journal of Economic Theory
JIE	Journal of International Economics
JMCB	Journal of Money, Credit and Banking
JME	Journal of Monetary Economics
JPE	Journal of Political Economy
JPubE	Journal of Public Economics
REStud	Review of Economic Studies
REStat	Review of Economics and Statistics
QJE	Quaterly Journal of Economics

A.3 Further Information of MCMC outputs

Table A.3: Various Estimates of the Parameters in the Probit Model – Simulation Data

	True values	MLE (known Y)	MCMC estimates (known Y)	MCMC estimates (unknown Y)
	(1)	(2)	(3)	**(4)**
A: Parameters of the covariates				
$\boldsymbol{\beta_1}$	0.8	0.788	0.814	**0.868**
		(0.019)	(0.018)	**(0.029)**
$\boldsymbol{\beta_2}$	-0.8	-0.776	-0.778	**-0.843**
		(0.028)	(0.029)	**(0.058)**
$\boldsymbol{\beta_3}$	1.4	1.287	1.305	**1.406**
		(0.059)	(0.063)	**(0.131)**
$\boldsymbol{\beta_4}$	-1.4	-1.295	-1.314	**-1.517**
		(0.056)	(0.037)	**(0.110)**
B: Cutoff points				
$\boldsymbol{c_2}$	1.769	1.794	1.901	**1.981**
		(0.068)	(0.064)	**(0.073)**
$\boldsymbol{c_3}$	4.125	4.010	4.142	**4.376**
		(0.089)	(0.023)	**(0.131)**

Note.—Column (1) presents true values in the data generating process. Columns (2)-(4) report the estimated probit paramters based on simulation data: column (2) presents the frequentist maximum likelihood estimates and column (3) reports the Bayesian MCMC estimates using the Metropolis-Hasting sampler given known outcomes. Column (4) is based on the proposed algorithm in which the outcome is treated as unknown. MLE standard errors or MCMC posterior standard deviations are reported in parentheses.

Figure A.1: Various Plots of Sampled Model Parameters from 5000 MCMC Iterations of the Proposed Algorithm – Simulation Data

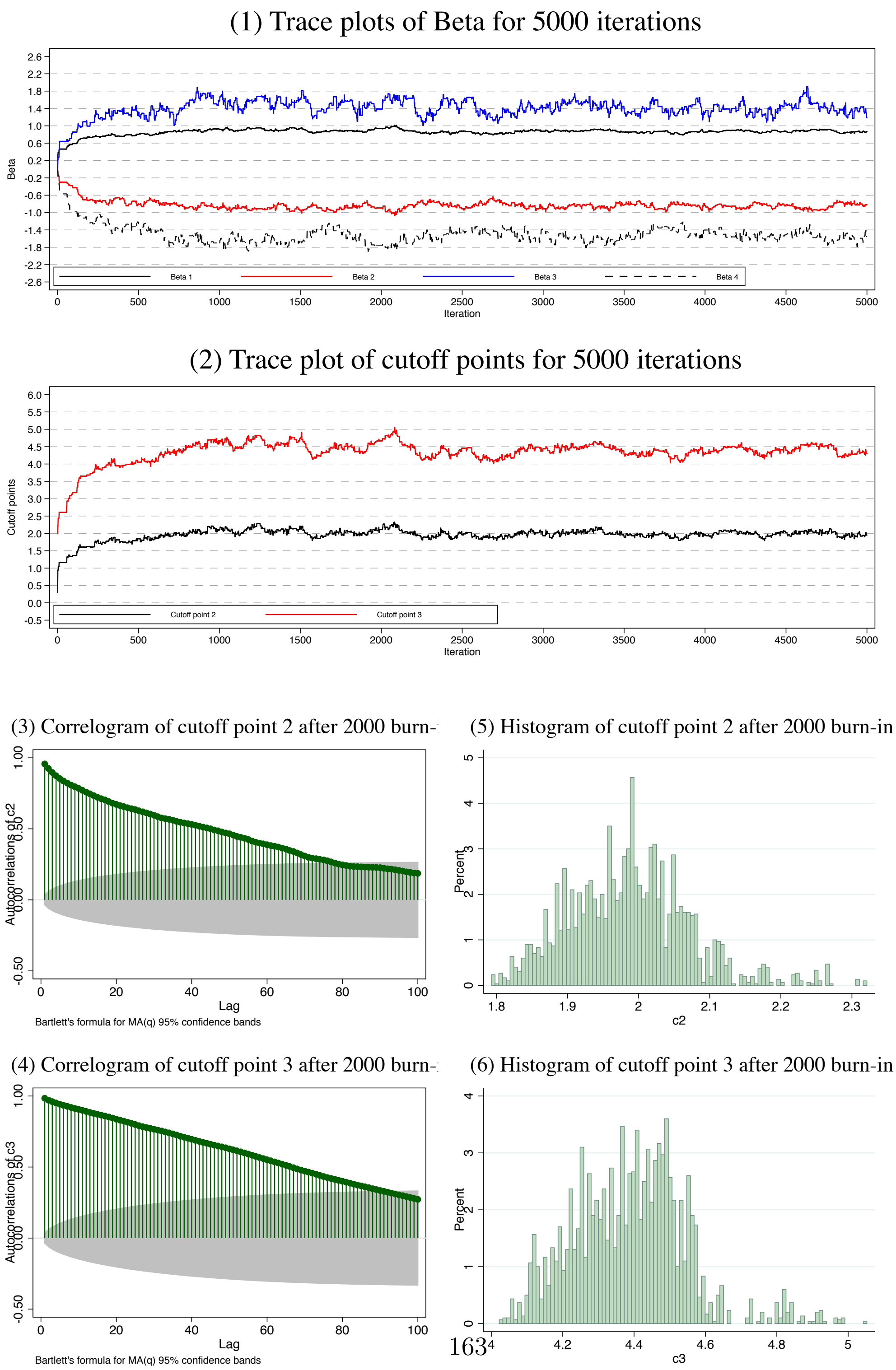


Figure A.2: VARIOUS PLOTS OF SAMPLED CUT-OFF POINTS FROM 20,000 MCMC ITERATIONS OF THE PROPOSED ALGORITHM (JOURNAL DUMMIES WITH MORE THAN 20 SUBMISSIONS TO THE REF ECONOMICS AND ECONOMETRICS PANEL)

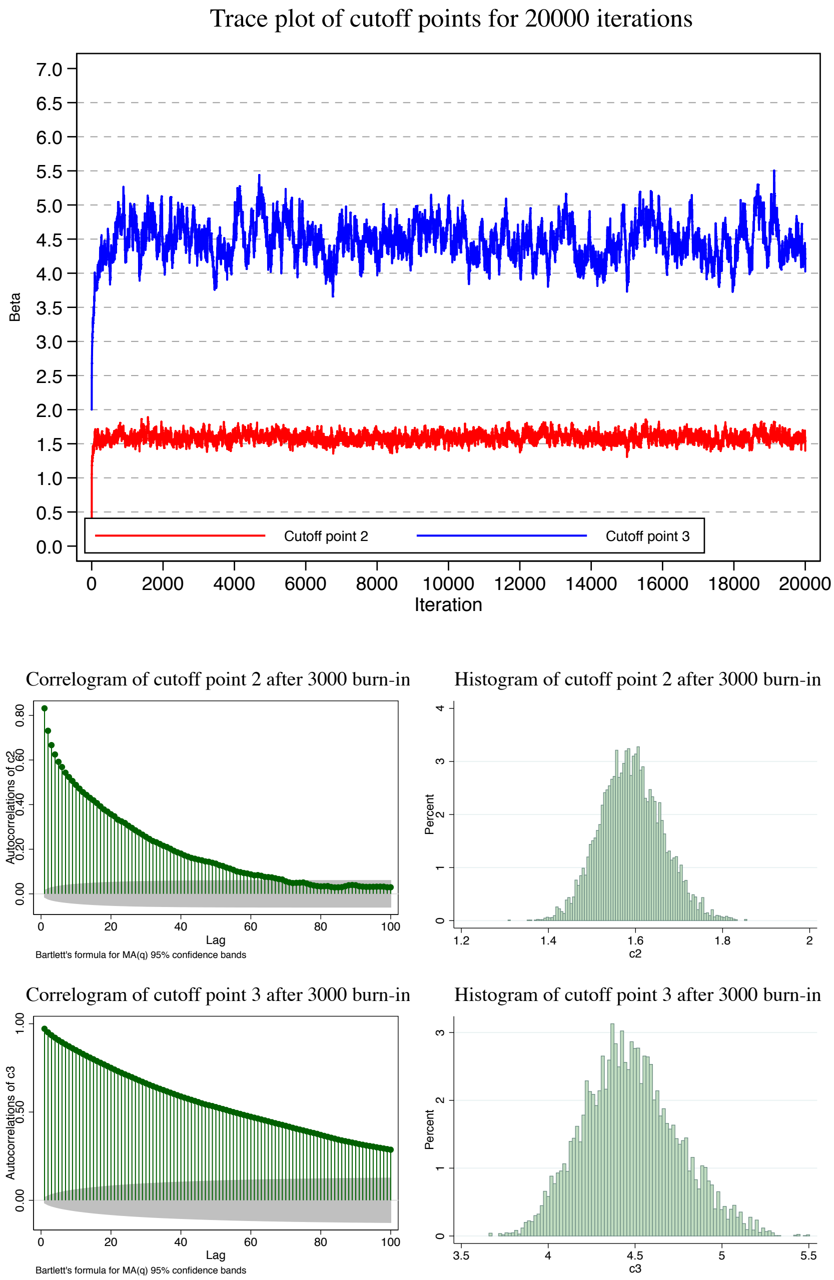


Figure A.3: Various Plots of Sampled Cut-Off Points from 20,000 MCMC Iterations of the Proposed Algorithm (Journal Dummies with More than 10 Submissions to the REF Economics and Econometrics Panel)

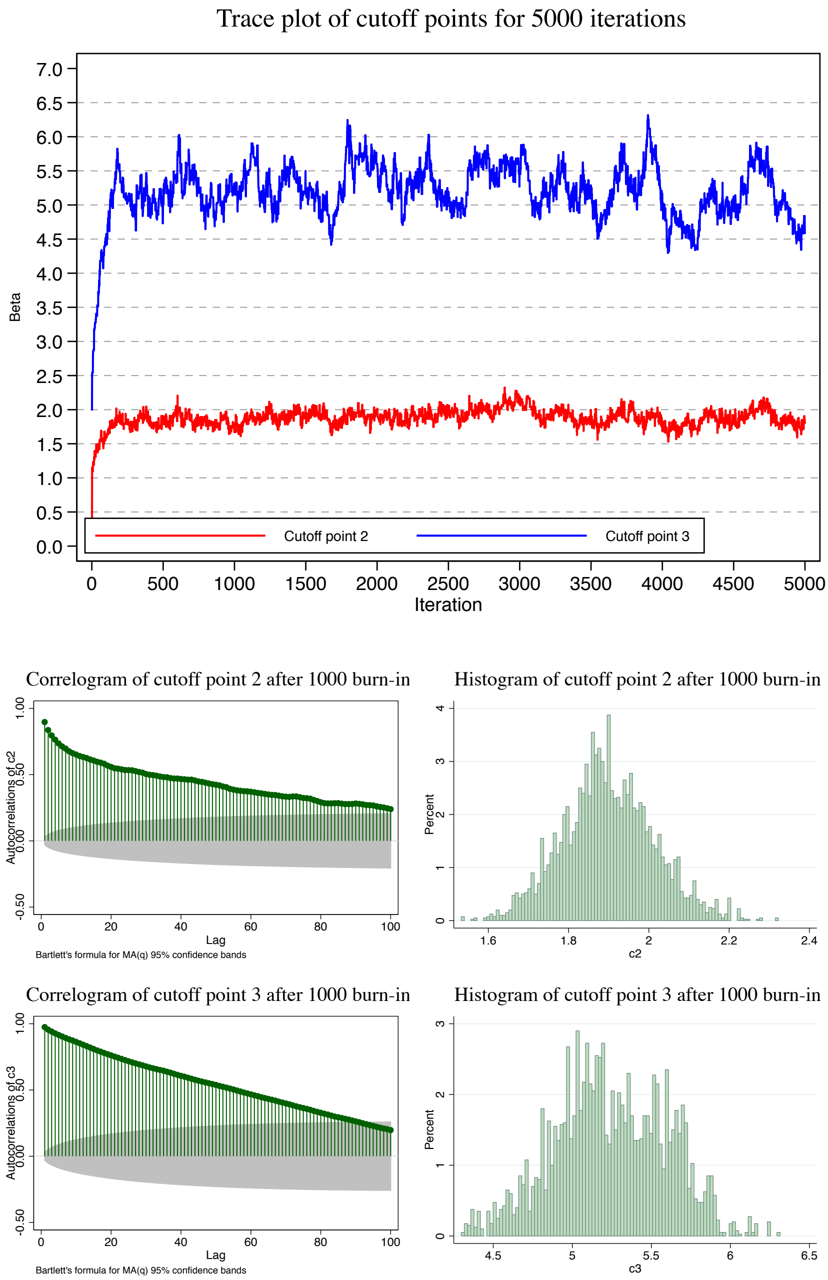


Figure A.4: Various Plots of Sampled Cut-Off Points from 20,000 MCMC Iterations of the Proposed Algorithm (Journal Dummies with More than 10 Submissions to the REF Business and Management Studies Panel)

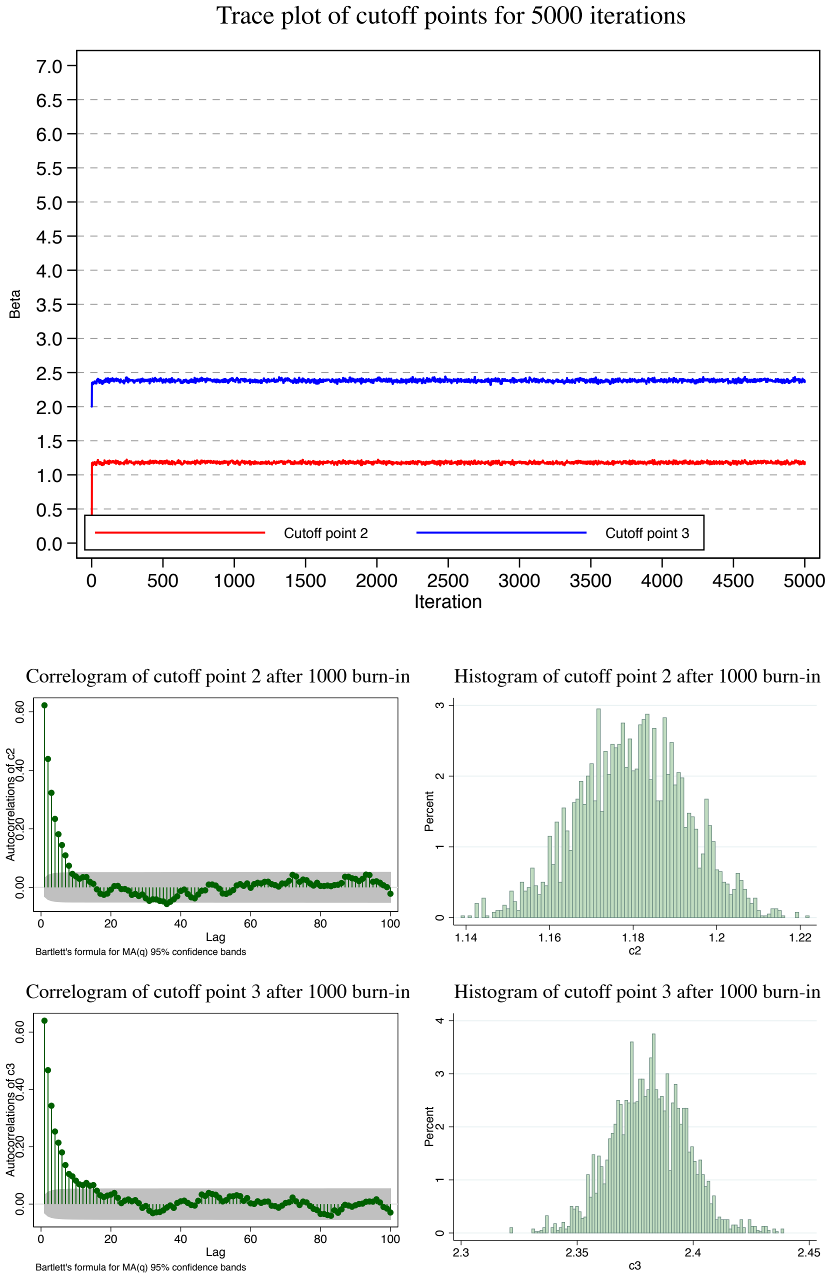
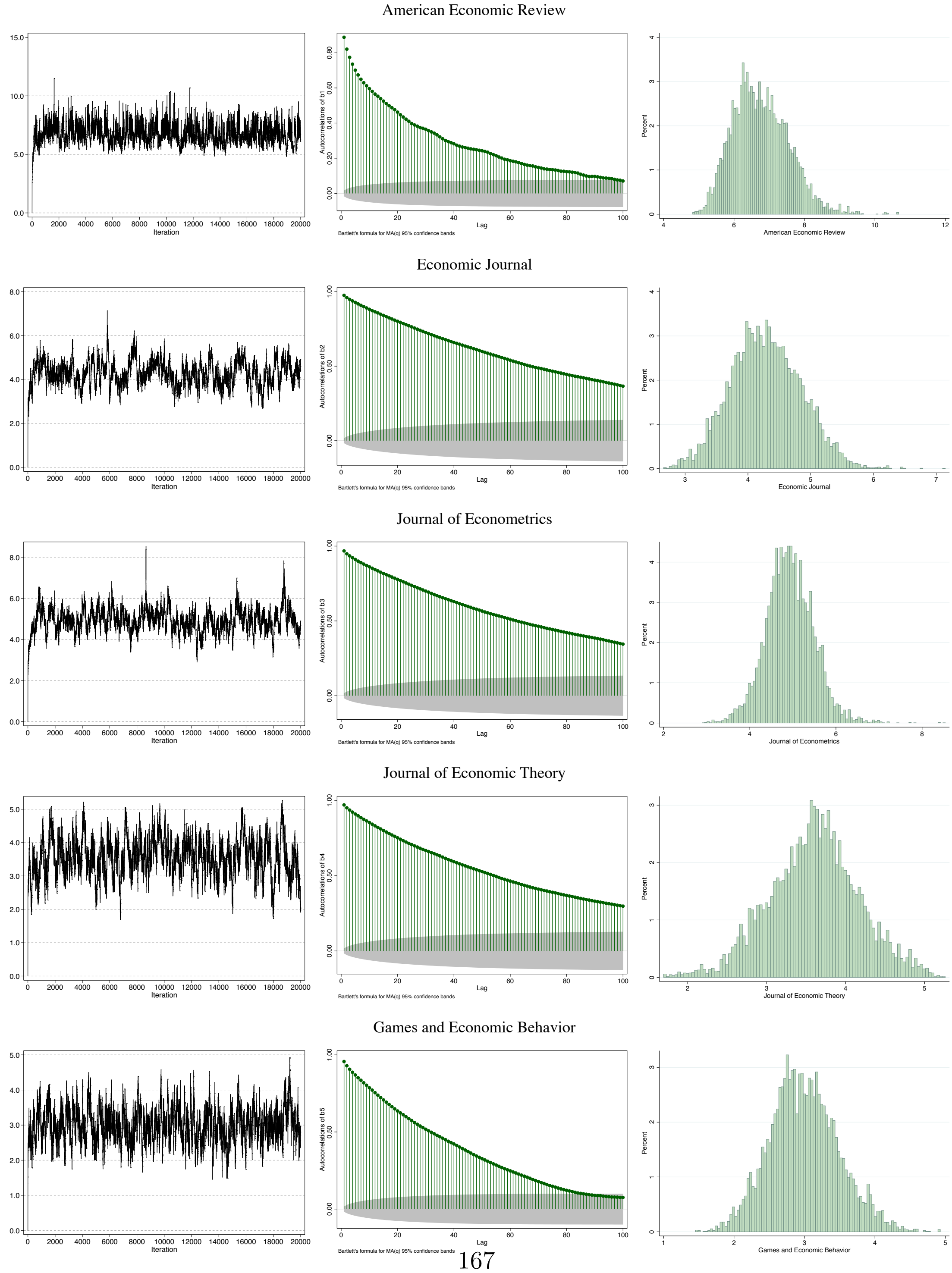
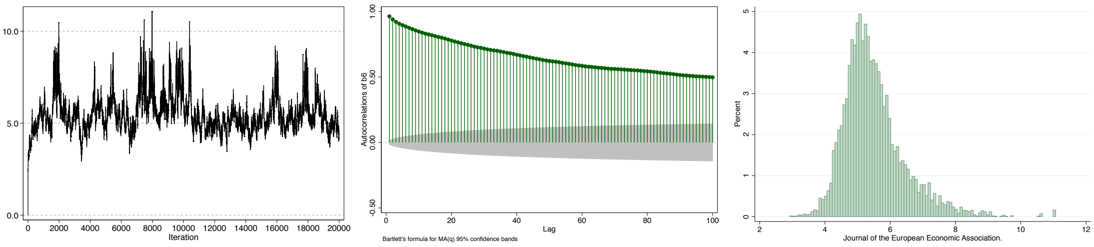


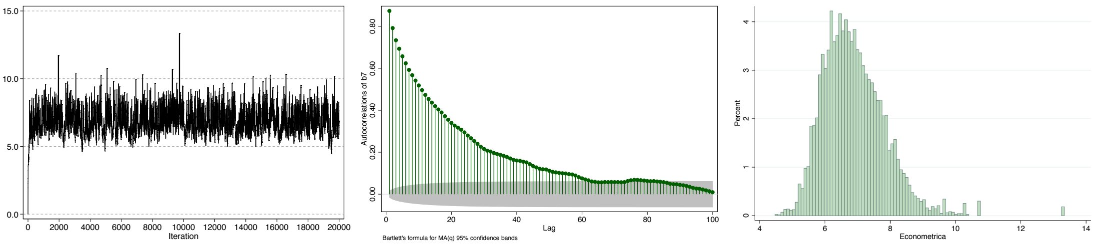
Figure A.5: Trace Plots (*left*), Correlograms (*middle*) and Histograms (*right*) of Sampled Parameters of Journal Dummies from 20,000 MCMC Iterations of the Proposed Algorithm (Journal Dummies with More than 20 Submissions to the REF Economics and Econometrics Panel)



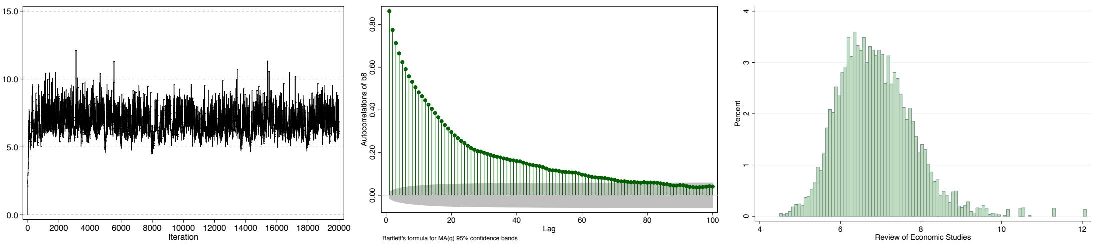
Journal of the European Economic Association.



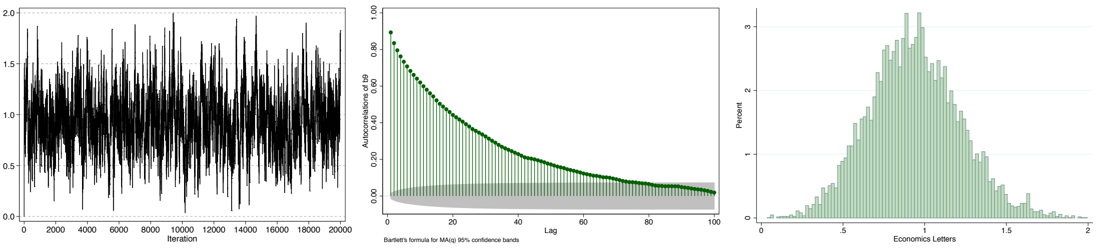
Econometrica



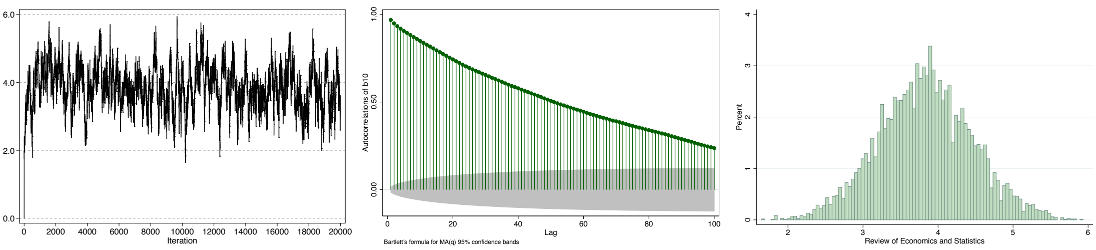
Review of Economic Studies



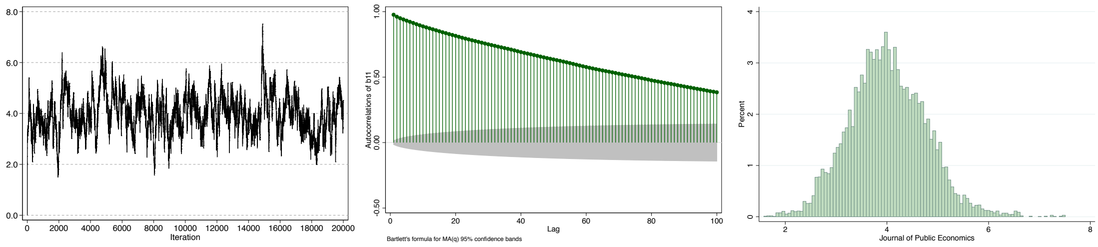
Economics Letters



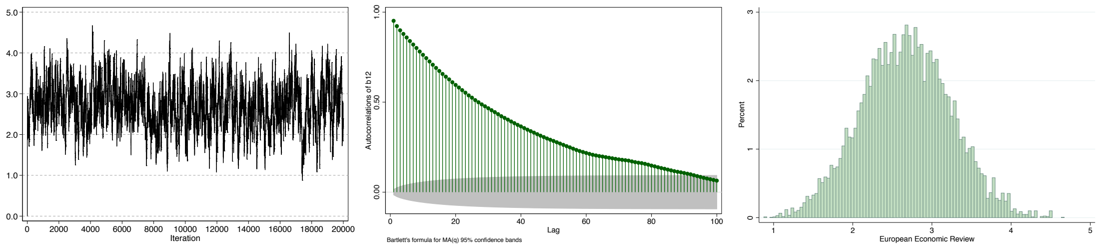
Review of Economics and Statistics



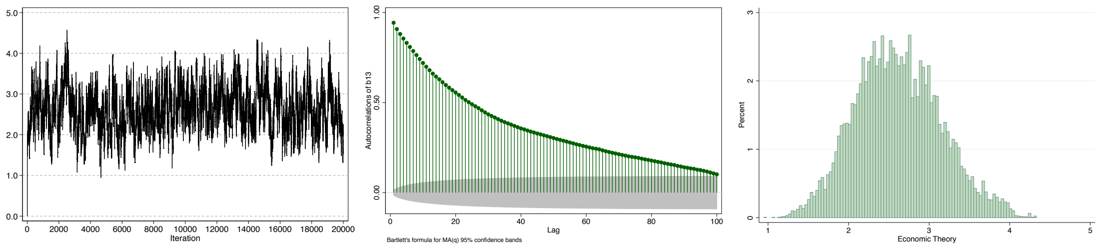
Journal of Public Economics



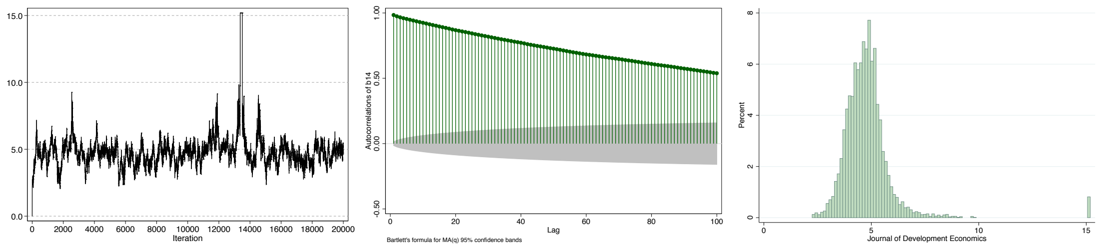
European Economic Review



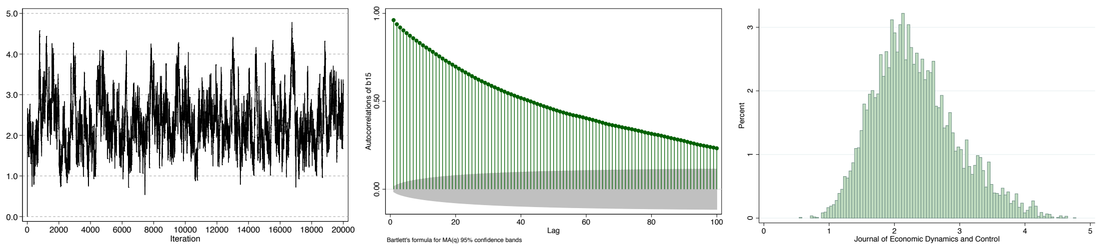
Economic Theory



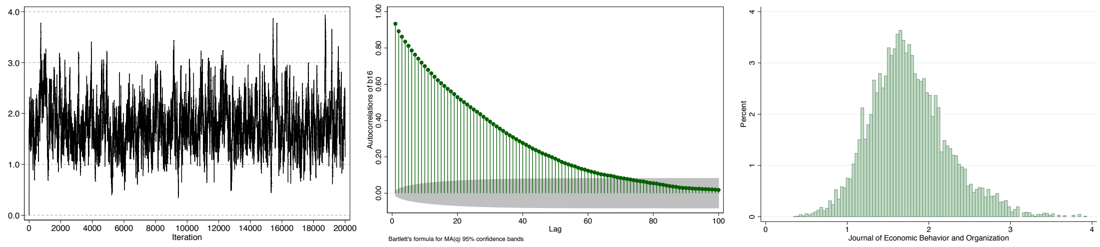
Journal of Development Economics



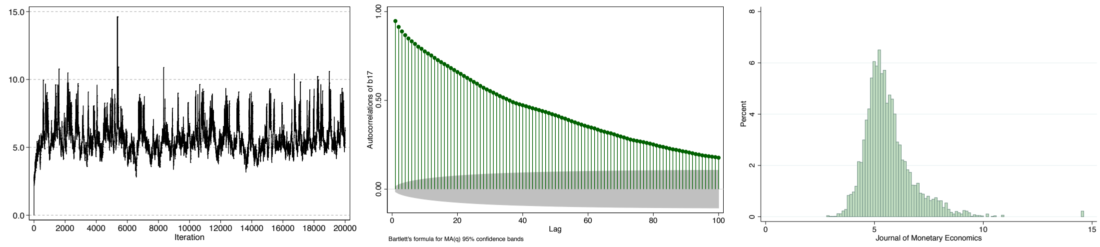
Journal of Economic Dynamics and Control



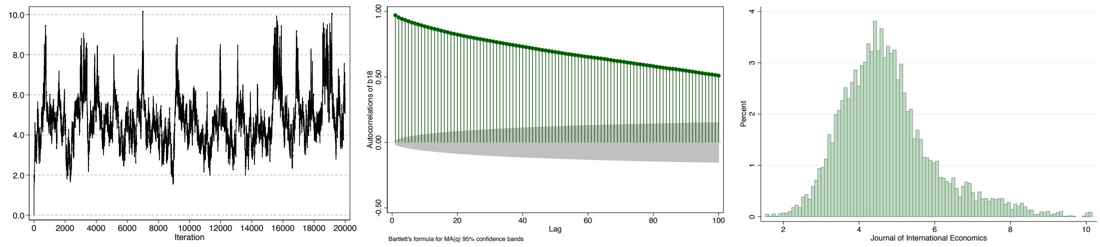
Journal of Economic Behavior and Organization



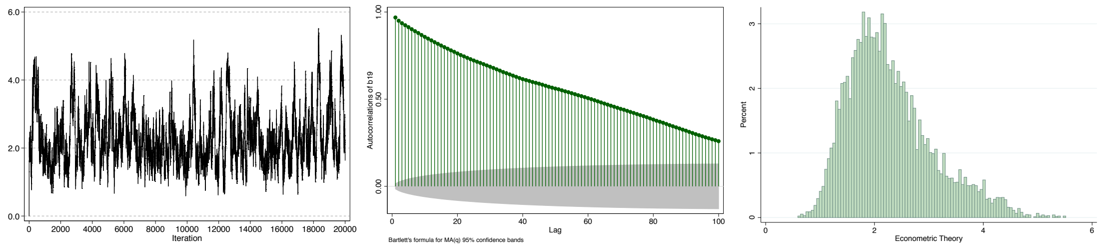
Journal of Monetary Economics



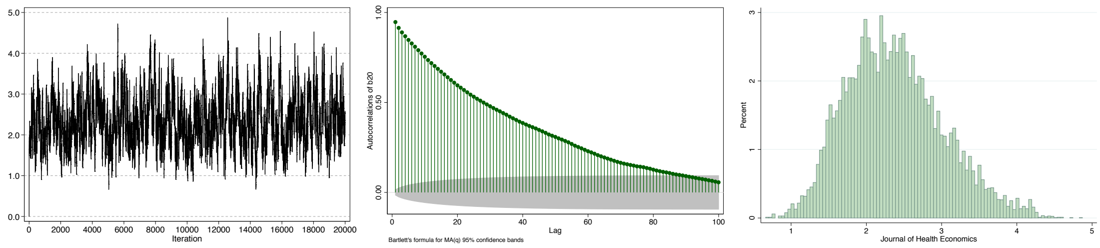
Journal of International Economics



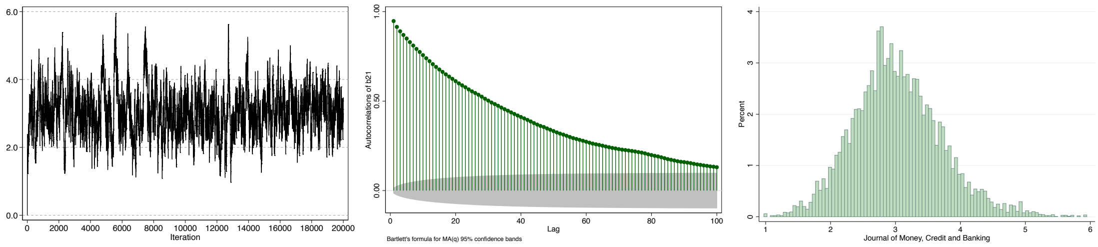
Econometric Theory



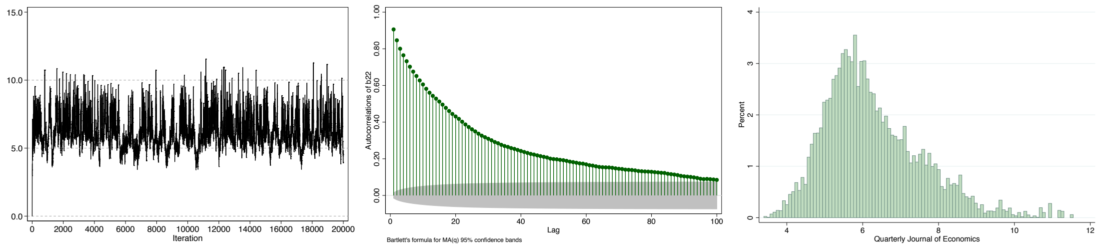
Journal of Health Economics



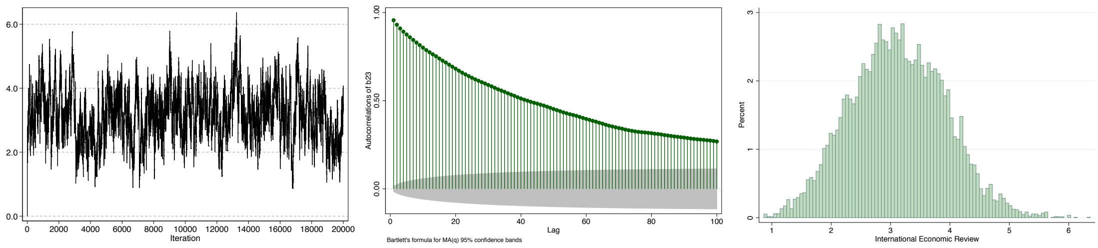
Journal of Money, Credit and Banking



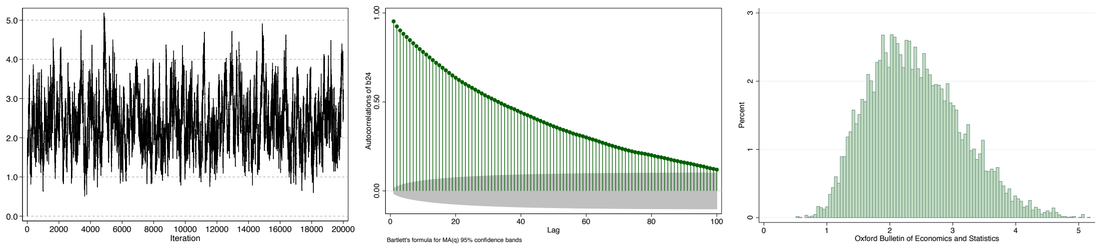
Quarterly Journal of Economics



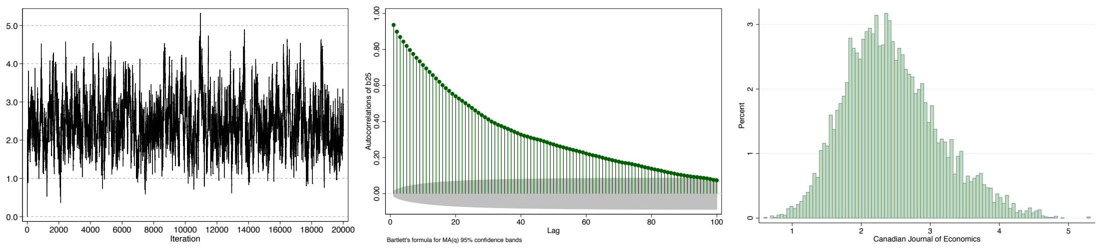
International Economic Review



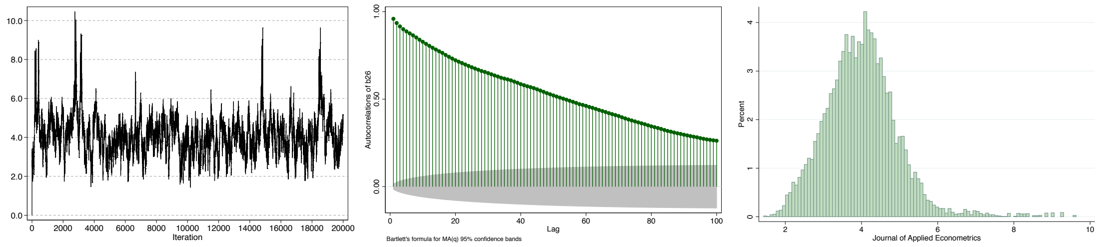
Oxford Bulletin of Economics and Statistics



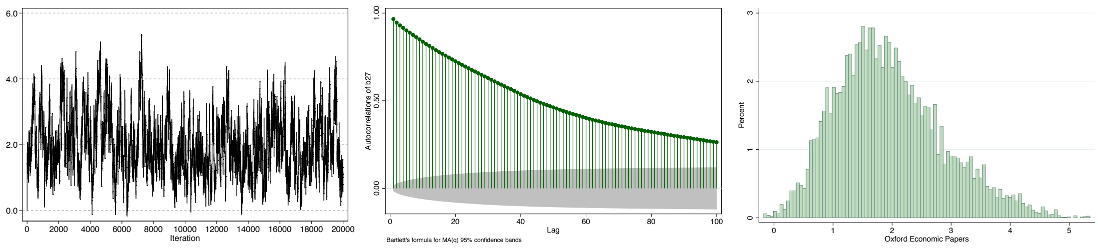
Canadian Journal of Economics



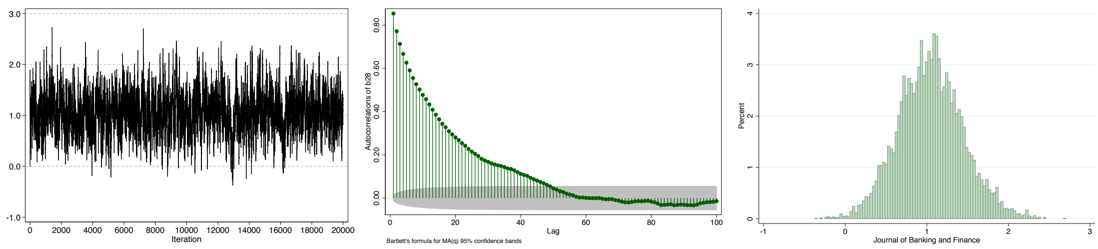
Journal of Applied Econometrics



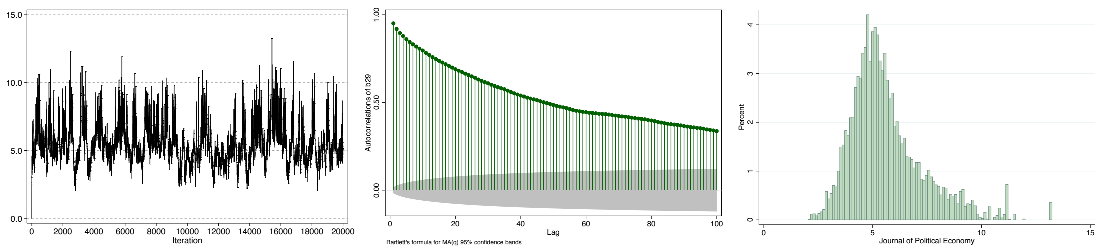
Oxford Economic Papers



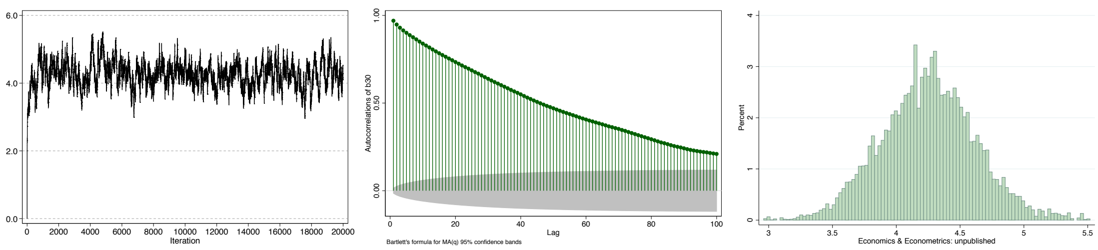
Journal of Banking and Finance



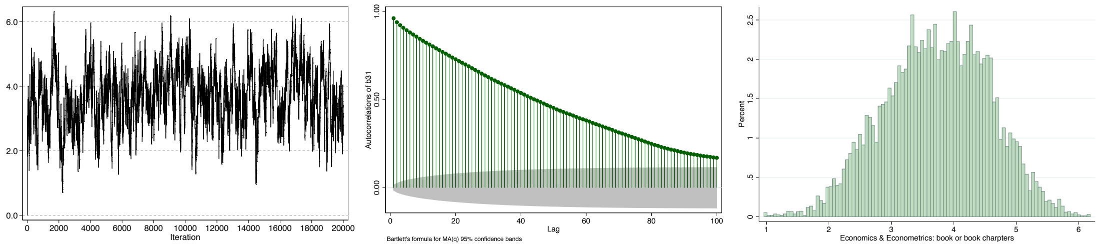
Journal of Political Economy



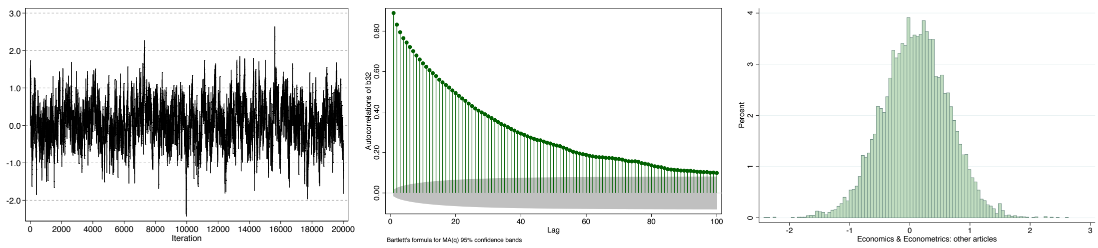
Economics & Econometrics: unpublished



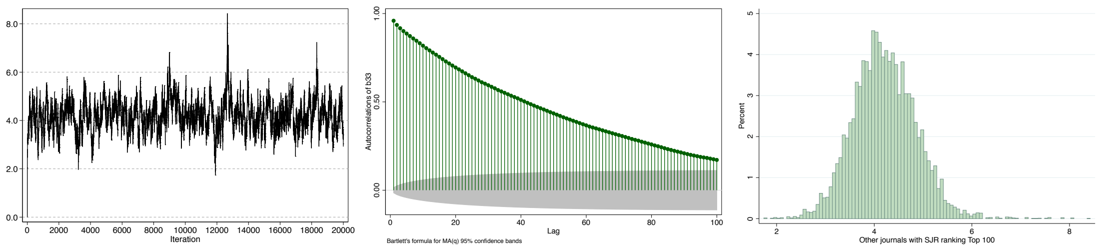
Economics & Econometrics: book or book charpters



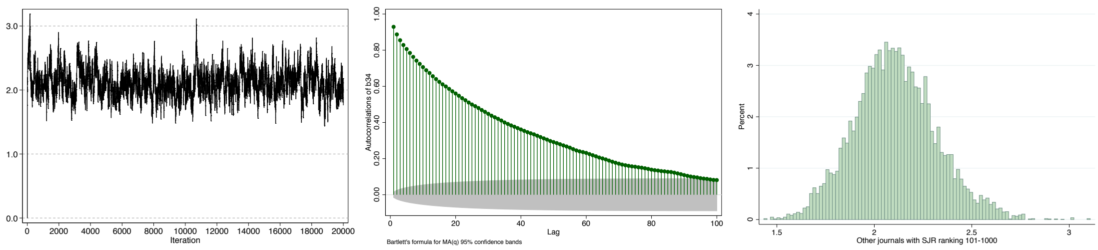
Economics & Econometrics: other articles



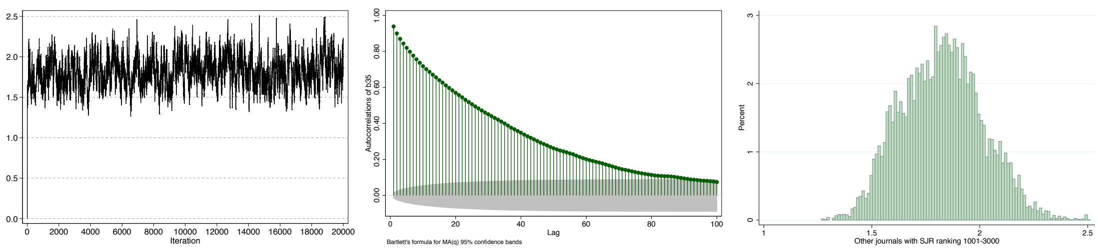
Other journals with SJR ranking Top 100



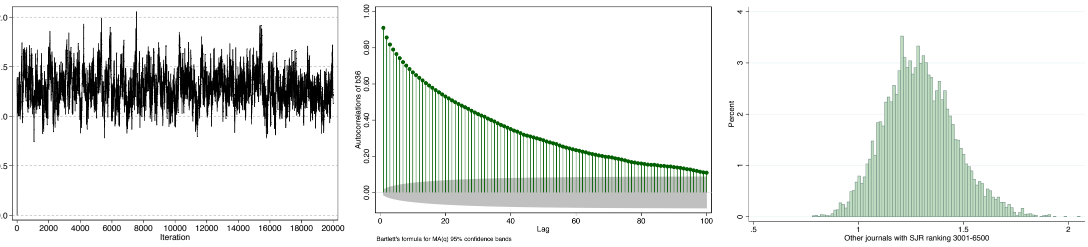
Other journals with SJR ranking 101-1000



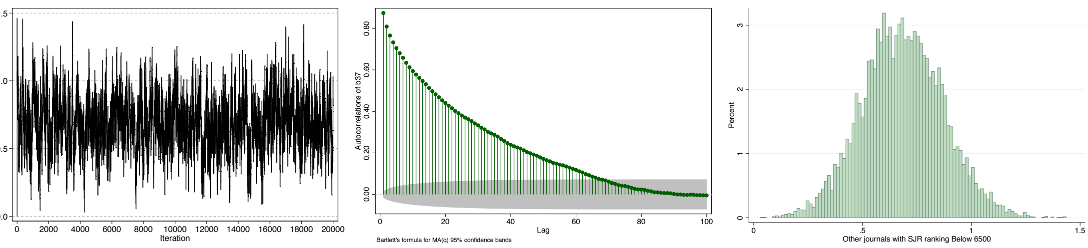
Other journals with SJR ranking 1001-3000



Other journals with SJR ranking 3001-6500



Other journals with SJR ranking Below 6500



B Appendices to Chapter 2

B.1 Applications to the Artificial Data

B.1.1 Exercise 1

In the first exercise, 500 observations used. (X_1, X_2, X_3, X_4) are generated from three clusters, X_4 is then transformed to a dummy variable by $X_4 = 1(X_4 > 0)$. Then we assign the treatment non-randomly based on a complicated relationship between covariates X_i and the binary intake T_i. In particular, the assignment of treatment follows a student t's distribution:

$$\begin{aligned}
\Pr(T = 1|X) &= \mathsf{t}_3(2 - 0.05(3(X_1 - 0.5))^3 + 1.2\sin(\frac{\pi X_2}{2}) \\
&\quad -2(\frac{0.05X_3 - 10}{10})^4 + 2X_4) \\
\mathcal{T}_i &= 1 \text{ if } U[0,1] > \Pr(T_i = 1|X)
\end{aligned}$$

and the potential outcomes are calculated by:

$$\begin{aligned}
\text{Control: } Y_i(0) &= 0.25 + 0.3X_{1i} + 0.4X_{2i} + 1.8X_{3i} - 0.4X_{4i} + \mathcal{N}(0,1) \\
\text{Treatment: } Y_i(1) &= y_{0i} + 2 + 1.2X_{1i} + 1.4X_{2i} + 2.8X_{3i} + 0.4X_{4i} + \mathcal{N}(0,1)
\end{aligned}$$

Hence, the true value ATT is $E(Y_i(1)|T_i = 1) - E(Y_i(0)|T_i = 1) = 3.7122$.

Figure B.1: Cross-plot of X_1 and X_2 of the True Data (artificial data 1)

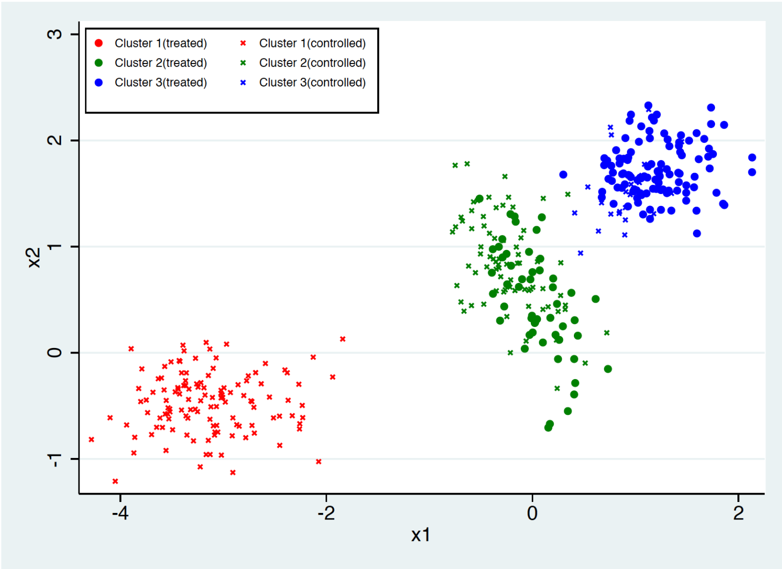


Figure B.2: Cross-plot of X_1 and X_2 in 200^{th} MCMC Iteration (artificial data 1)

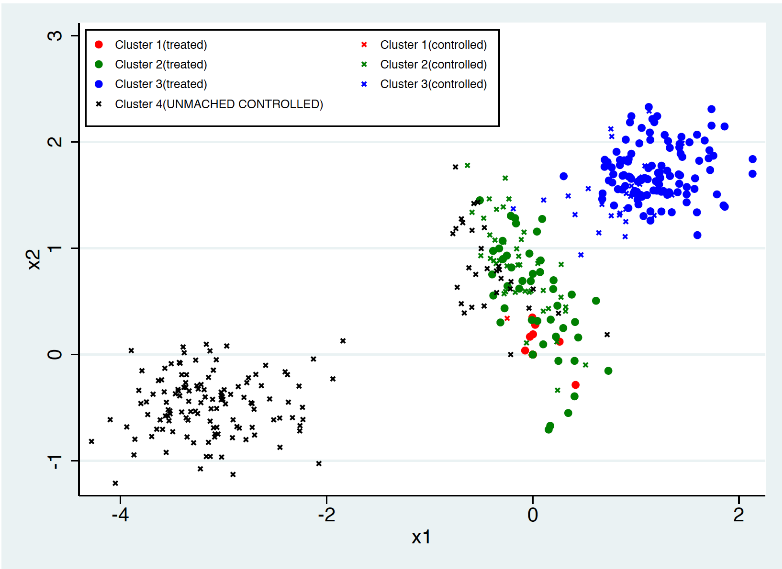


Table B.1: Estimated Results using the Artificial Data 2 (true effect 3.7122).

	Estimated ATT	SD/SE
DP matching	3.6802	0.556
Radius matching via p-score	2.294	0.274
Stratification matching via p-score	3.646	.
Nearest Neighbour matching via p-score	3.853	0.226
Kernel matching via p-score	3.653	0.222

Notes: The first row presents the MCMC posterior mean and standard deviation of ATT estimator based on DP matching. For the radius matching, the size of radius is 0.1 by default of the Stata *pscore* program (Becker and Ichino, 2002). Kernel matching uses the Gaussian kernel with bandwidth 0.06 by default.

B.1.2 Exercise 2

This synthetic data is generated to mimic an observational study in which part of the control group possess different distributions of covariates X_i. Firstly, 500 observations are generated from three clusters and randomly divided into treatment and controlled groups. Secondly, 450 additional 'noisy' controlled observations are constructed on the basis of two different distributions. These additional control groups are supposed to be eliminated by matching algorithms. Finally, their outcomes are defined by:

$$\begin{aligned} \text{Control: } Y_i(0) &= 0.25 + 1.4X_{1i} + 2.5X_{2i} + 3.8X_{3i} + \mathcal{N}(0,1) \\ \text{Treatment: } Y_i(1) &= Y_i(0) + 1.5 + 3X_{1i} + 2.4X_{2i} + 2.8X_{3i} + \mathcal{N}(0,1) \end{aligned}$$

hence, the true value of ATT is $E(Y_i(1)|T_i = 1) - E(Y_i(0)|T_i = 1) = 1.6633$.

Figure B.3: Cross-plot of X_1 and X_2 of the True Data (artificial data 2)

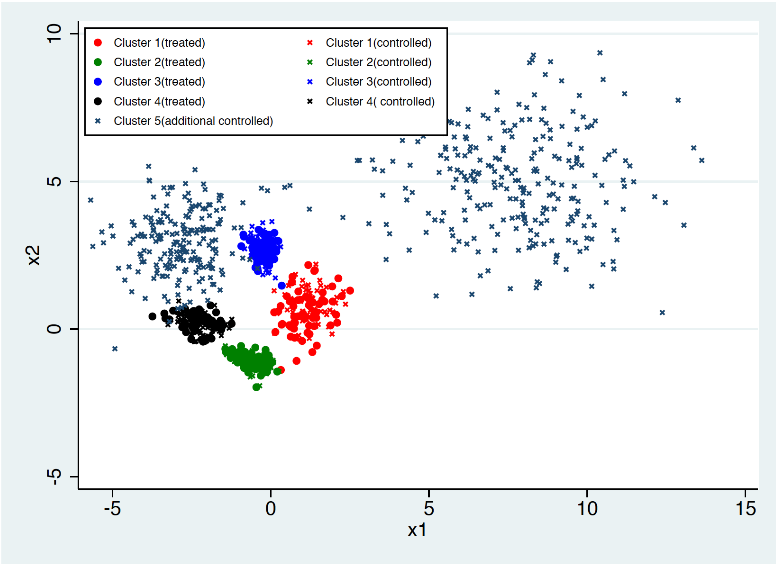


Figure B.4: Cross-plot of X_1 and X_2 in 200^{th} MCMC Iteration (artificial data 2)

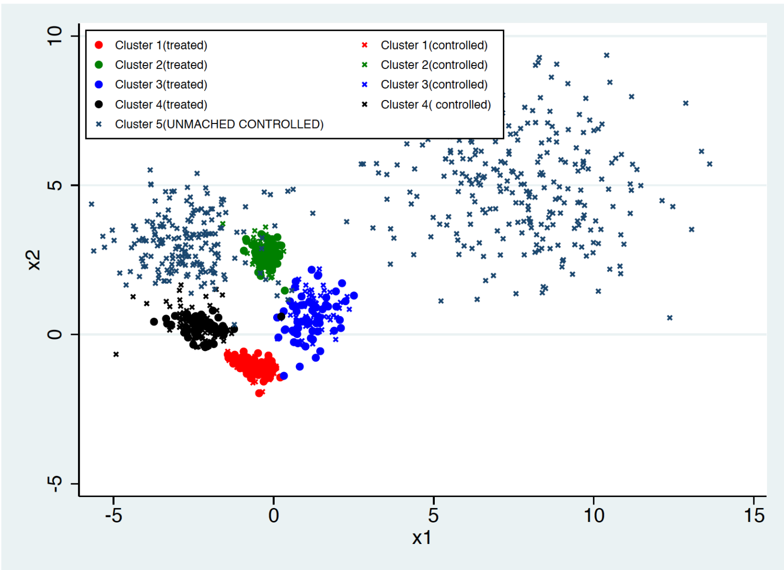


Table B.2: Estimated Results using the Artificial Data 2 (true effect 1.6633).

	Estimated ATT	SD/SE
DP matching	1.4809	0.230
Radius matching via p-score	3.797	1.190
Stratification matching via p-score	2.227	0.457
Nearest Neighbour matching via p-score	2.262	0.678
Kernel matching via p-score	1.254	0.373

Notes: The first row presents the MCMC posterior mean and standard deviation of ATT estimator based on DP matching . For the radius matching, the size of radius is 0.1 by default of the Stata *pscore* program (Becker and Ichino, 2002). Kernel matching uses the Gaussian kernel with bandwidth 0.06 by default.

B.2 Visualizations of the Metaphor of Chinese Restaurant with Invited Guests

Figure B.5: Chinese Restaurant Process for the Treatment Group

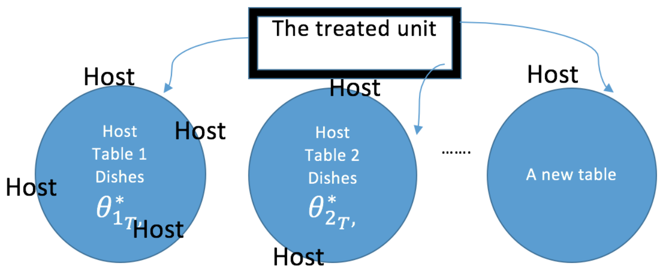
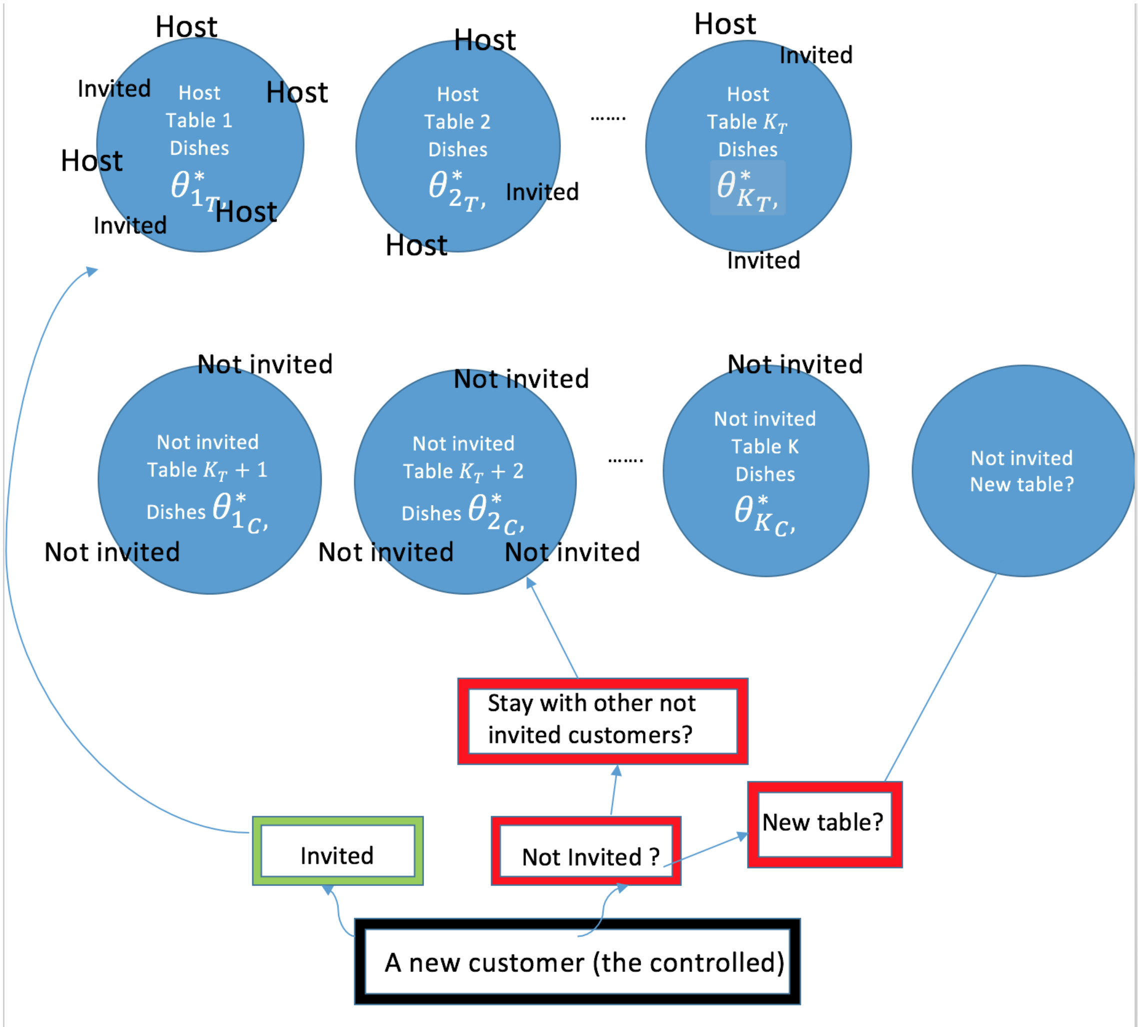


Figure B.6: Chinese Restaurant with Invited Guests



C Appendices to Chapter 3

Table C.1: Fathers' Characteristics before Treatment (2012)

	Treatment mean	Control mean	Difference in means	Std. error
	(1)	(2)	(3)	(4)
Age	35.053	33.645	1.408	(0.979)
From Quito	0.506	0.513	-0.007	(0.061)
Highest educational level: primary school	0.446	0.505	-0.059	(0.063)
Highest educational level: secondary school	0.516	0.466	0.050	(0.064)
Highest educational level: university	0.038	0.029	0.009	(0.023)
Not religious	0.118	0.094	0.023	(0.039)
Christian	0.843	0.858	-0.015	(0.045)
Worked	0.873	0.870	0.004	(0.041)
Worked full-time	0.938	0.880	0.058	(0.036)
Self-employed	0.828	0.838	-0.011	(0.049)
Worked in the formal sector	0.375	0.460	-0.085	(0.064)
Mean firm size	26.693	42.388	-15.695	(10.792)
$F(10, 198) = 1.3881$ $Prob > F = 0.1879$				
Observations	166	115	281	

Note. The numbers attached in columns 1-3 of the last row of the table indicate the numbers of observations in the treated sample, control sample and the total sample, respectively. Statistics are based on the 2012 survey of parents. Standard errors are presented in parentheses in the column (4); * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. An F-test on the overall significance of the pre- treatment variables is shown at the end of the table.

Table C.2: Household Characteristics before Treatment (2012)

	Treatment mean (1)	Control mean (2)	Difference in means (3)	Std. error (4)
Family lived in Pisulli	0.675	0.583	0.092	(0.058)
House was owned	0.285	0.122	0.163***	(0.049)
House had drinkable	0.770	0.878	-0.109**	(0.047)
House had electricity	0.970	0.991	-0.022	(0.018)
House had toilet inside	0.430	0.383	0.048	(0.060)
Average number of rooms	3.667	3.209	0.458**	(0.218)
Family who had no vehicles	0.946	0.913	0.033	(0.031)
Family who had bicycles	0.024	0.052	-0.028	(0.022)
Family who had other means of transport	0.030	0.035	-0.005	(0.021)
Family average monthly wage (USD)	248.788	247.807	0.981	(17.250)
$F(9, 286) = 3.0513$ Prob > F = 0.0017				
Observations	166	115	281	

Note. The numbers attached in columns 1-3 of the last row of the table indicate the numbers of observations in the treated sample, control sample and the total sample, respectively. Statistics are based on the 2012 survey. Standard errors are presented in parentheses in the column (4); * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. An F-test on the overall significance of the pre- treatment variables is shown at the end of the table.

Table C.3: Household Characteristics before Treatment (2012)

	Treatment mean	Control mean	Difference in means	Std. error
	(1)	(2)	(3)	(4)
A: Child characteristics				
Age	8.633	8.633	0.000	(0.273)
Birth order	1.819	1.819	0.000	(0.274)
Number of children mother had in 2005	1.849	1.849	0.000	(0.326)
Number of young siblings in 2005	0.307	0.307	0.000	(0.094)
B: Mother characteristics				
Age	31.988	31.989	-0.001	(1.035)
Worked	0.470	0.470	0.000	(0.082)
Worked full-time	0.259	0.259	0.000	(0.066)
Mean firm size	10.120	10.123	-0.002	(6.924)
Single before	0.181	0.181	0.000	(0.063)
From Quito	0.560	0.560	0.000	(0.082)
Did not complete primary	0.114	0.121	-0.006	(0.053)
Completed primary	0.392	0.392	0.000	(0.083)
Did not complete secondary	0.295	0.295	0.000	(0.072)
Completed secondary	0.169	0.169	0.000	(0.060)
Started university	0.024	0.024	0.000	(0.027)
C: Father characteristics				
Age	33.873	33.874	-0.001	(1.396)
Worked before	0.873	0.874	0.000	(0.057)
Mean firm size	17.373	17.394	-0.020	(8.161)
Did not complete primary	0.072	0.072	0.000	(0.037)
Completed primary	0.349	0.349	0.000	(0.077)
Did not complete secondary	0.355	0.355	0.000	(0.086)
Completed secondary	0.133	0.133	0.000	(0.051)
Started university	0.036	0.036	0.000	(0.027)
D: Household characteristics				
Parents were married	0.458	0.458	0.000	(0.084)
Parents cohabited	0.343	0.343	0.000	(0.076)
Parents from the same city	0.470	0.470	0.000	(0.083)
Family monthly wage	247.289	247.302	-0.013	(23.870)
Observations	166	115	281	

Note. The numbers attached in columns 1-3 of the last row of the table indicate the numbers of observations in the treated sample, control sample and total sample, respectively. Statistics are based on the 2012 survey. Standard errors are presented in parentheses in the column (4); * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. An F-test on the overall significance of the pre- treatment variables is shown at the end of the table.

Table C.4: ESTIMATED EFFECTS ON FATHERS' LABOR MARKET OUTCOMES (2012)

	Not weighted			Weighted		
	(1)	(2)	(3)	(4)	(5)	(6)
Summary index	–0.110	–0.126	–0.113	–0.084	–0.080	–0.097
	(0.103)	(0.104)	(0.105)	(0.128)	(0.108)	(0.105)
Working	–0.017	–0.017	–0.021	–0.023	–0.023	–0.024
	(0.025)	(0.028)	(0.029)	(0.025)	(0.024)	(0.024)
Working full-time	–0.030	–0.034	–0.029	–0.006	–0.006	–0.012
	(0.043)	(0.045)	(0.046)	(0.065)	(0.048)	(0.046)
Working with contract	–0.070	–0.085	–0.063	–0.045	–0.040	–0.051
	(0.065)	(0.070)	(0.071)	(0.089)	(0.084)	(0.082)
Child Controls	No	Yes	Yes	No	Yes	Yes
Household Demographics	No	Yes	Yes	No	Yes	Yes
Household Economics	No	No	Yes	No	No	Yes
Observations	281	281	281	281	281	281

NOTE. Each cell reports the estimated treatment effect from a separate regression based on the 2012 survey. Estimated summary indices of corresponding outcomes are reported in shading rows. Column (1)-(3) present results using the original sample without entropy balancing. Column (4)-(6) stem from the weighted sample adjusted by entropy balancing. Standard errors are presented in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C.5: ESTIMATED EFFECTS ON SELF-ESTEEM, BIG FIVE PERSONALITY TRAITS AND FERTILITY CHOICES (2012)

	Not weighted			Weighted		
	(1)	(2)	(3)	(4)	(5)	(6)
Summary index†	0.038	0.033	0.045	0.060	0.060	0.063
	(0.065)	(0.067)	(0.068)	(0.083)	(0.090)	(0.083)
Rosenberg scale	0.034	0.019	0.017	–0.006	–0.004	–0.001
	(0.056)	(0.058)	(0.060)	(0.077)	(0.060)	(0.059)
Agreeableness	–0.015	–0.023	–0.017	0.015	0.017	0.017
	(0.067)	(0.070)	(0.072)	(0.101)	(0.088)	(0.082)
Conscientiousness	0.010	–0.014	0.002	0.019	0.018	0.019
	(0.078)	(0.083)	(0.086)	(0.109)	(0.095)	(0.091)
Extraversion	–0.080	–0.062	–0.028	–0.048	–0.053	–0.053
	(0.066)	(0.071)	(0.073)	(0.097)	(0.087)	(0.085)
Neuroticism	0.080	0.092	0.088	0.078	0.082	0.083
	(0.070)	(0.075)	(0.077)	(0.093)	(0.084)	(0.082)
Openness to experience	0.063	0.082	0.091	0.128	0.129	0.130
	(0.075)	(0.081)	(0.083)	(0.117)	(0.103)	(0.100)
Pregnant	–0.017	–0.016	–0.018	–0.043	–0.042	–0.041
	(0.019)	(0.021)	(0.022)	(0.044)	(0.038)	(0.038)
More children (including pregnant women)?	0.068	0.066	0.063	0.011	0.012	0.010
	(0.048)	(0.051)	(0.053)	(0.074)	(0.064)	(0.062)
Child Controls	No	Yes	Yes	No	Yes	Yes
Household Demographics	No	Yes	Yes	No	Yes	Yes
Household Economics	No	No	Yes	No	No	Yes
Observations	281	281	281	281	281	281

NOTE. Each cell reports the estimated treatment effect from a separate regression based on the 2012 survey. Estimated summary indices of corresponding outcomes are reported in shading rows. Column (1)-(3) present results using the original sample without entropy balancing. Column (4)-(6) stem from the weighted sample adjusted by entropy balancing. Standard errors are presented in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

† This summary index is constructed by only outcomes of Rosenberg self-esteem scale and Big Five Personality Traits.

Table C.6: Estimated Effects on Fathers' Labor Market Outcomes (Pooled data of 2012 and 2013)

	Not weighted			Weighted		
	(1)	(2)	(3)	(4)	(5)	(6)
Summary index	0.011	–0.026	–0.006	–0.075	–0.062	–0.065
	(0.083)	(0.081)	(0.078)	(0.107)	(0.089)	(0.084)
Working	0.006	–0.008	–0.006	–0.020	–0.015	–0.016
	(0.032)	(0.031)	(0.030)	(0.035)	(0.028)	(0.028)
Working full-time	0.032	0.018	0.023	–0.008	–0.002	–0.003
	(0.043)	(0.044)	(0.043)	(0.049)	(0.043)	(0.041)
Working with contract	–0.023	–0.045	–0.023	–0.057	–0.053	–0.054
	(0.053)	(0.060)	(0.058)	(0.080)	(0.068)	(0.065)
Child Controls	No	Yes	Yes	No	Yes	Yes
Household Demographics	No	Yes	Yes	No	Yes	Yes
Household Economics	No	No	Yes	No	No	Yes
Observations	496	496	496	496	496	496

Note. Each cell reports the estimated treatment effect from a separate regression based on the 2012 and 2013 survey. Estimated summary indices of corresponding outcomes are reported in shading rows. Column (1)-(3) present results using the original sample without entropy balancing. Column (4)-(6) stem from the weighted sample adjusted by entropy balancing. Standard errors are presented in parentheses and are clustered at the maternal level; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C.7: CHARACTERISTICS AND PRE-PROGRAM OUTCOMES OF THE TREATMENT SAMPLE AND THE SAMPLE THAT LEFT THE PROGRAM AND ENROLLED BETWEEN 2005 AND 2009

	Attrition mean	Attrition mean (≤4 years)	Treatment mean	Difference in means (1)-(3)	Difference in means (2)-(3)
	(1)	(2)	(3)	(4)	(5)
Mother age when enrolled	24.959	25.676	26.185	-1.226	-0.509
				(1.000)	(1.168)
Number of children in 2005	1.616	1.657	1.849	-0.233	-0.192
				(0.163)	(0.203)
Mother lived together with partner	0.804	0.789	0.801	0.002	-0.012
				(0.062)	(0.072)
Mother worked	0.412	0.530	0.470	-0.058	0.060
				(0.055)	(0.073)
Highest educational level: primary school	0.693	0.694	0.687	0.006	0.007
				(0.052)	(0.069)
Highest educational level: secondary school	0.216	0.242	0.169	0.047	0.073
				(0.044)	(0.058)
Highest educational level: started university	0.026	0.048	0.024	0.002	0.024
				(0.018)	(0.026)
Family lived in Pisulli	0.641	0.559	0.675	-0.034	-0.116*
				(0.052)	(0.069)
Child age when enrolled	2.680	2.575	2.696	-0.016	-0.122
				(0.157)	(0.198)
Observations of mothers	172	70	162		
Observations of children	258	111	219		

NOTE. The column (1) is is based on the sample of mothers/children who enrolled between 2005 and 2009 and left the program. The column (2) is based on the sample who enrolled between 2005 and 2009 and left in 4 years. The column (3) presents the treatment means based on the 2012 survey. Standard errors are presented in parentheses in the column (4); * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C.8: Estimated Effects of Number of Years in the Program on Mothers' Outcomes (2012)

	Not weighted		Weighted	
	(1)	(2)	(3)	(4)
Labor market outcomes				
Works	0.035***	0.029***	0.037***	0.034***
	(0.010)	(0.010)	(0.013)	(0.011)
Working full-time	0.038***	0.037***	0.040***	0.043***
	(0.009)	(0.010)	(0.010)	(0.009)
Working with contract	0.038***	0.039***	0.043***	0.042***
	(0.008)	(0.008)	(0.008)	(0.007)
Average family monthly income	7.794**	8.139**	7.871*	7.993**
	(3.509)	(3.411)	(4.288)	(3.564)
Mothers' economic and social independence				
Manage own money	0.038***	0.037***	0.036***	0. 037***
	(0.010)	(0.011)	(0. 014)	(0. 011)
Participates in voluntary activities	0.013	0.024**	0.012	0.018
	(0.011)	(0.012)	(0.015)	(0.012)
Currently in school	0.012**	0.017***	0.013**	0.014**
	(0.006)	(0.007)	(0.006)	(0.006)
Own or joint decision on own work status	0.022***	0.024***	0.021**	0.023**
	(0.007)	(0.007)	(0.010)	(0.009)
Household decisions- making				
Own/joint decision on child's education	0.017***	0.018***	0.024**	0.024**
	(0.006)	(0.007)	(0.010)	(0.009)
Own/joint decision on own health	0.002	0.001	-0.004	-0.004
	(0.005)	(0.005)	(0.004)	(0.005)
Own/joint decision on on discipline	0.013**	0.013*	0.018*	0.016*
	(0.007)	(0.007)	(0.010)	(0.009)
Own/joint decision on expenditure	0.014*	0.017*	0.023*	0.023**
	(0.008)	(0.009)	(0.013)	(0.010)
Own/joint decision on food expenditure	0.008	0.008	0.010	0.010
	(0.008)	(0.009)	(0.013)	(0.010)
Own/joint decision on having children	0.003	0.005	0.001	0.002
	(0.004)	(0.004)	(0.004)	(0.004)
Own/joint decision on contraceptives	-0.003	-0.001	-0.004	-0.004
	(0.005)	(0.006)	(0.007)	(0.005)
Child Controls	No	Yes	No	Yes
Household Demographics	No	Yes	No	Yes
Household Economics	No	Yes	No	Yes
Observations	281	281	281	281

Note. Each cell reports the estimated effect of years of treatment on the mothers' outcome from a separate regression based on the 2012 survey. Columns (1)-(2) present results using the original sample without entropy balancing. Columns (3)-(4) stem from the weighted sample adjusted by entropy balancing. Standard errors are presented in parentheses and are clustered at the maternal level; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C.9: Estimated Effects of Number of Years in the Program on Mothers' Outcomes (Pooled 2012 and 2013)

	Not weighted		Weighted	
	(1)	(2)	(3)	(4)
Labor market outcomes				
Works	0.035***	0.026**	0.033**	0.031***
	(0.009)	(0.010)	(0.013)	(0.011)
Working full-time	0.035***	0.033***	0.039***	0.041***
	(0.009)	(0.009)	(0.009)	(0.009)
Working with contract	0.033***	0.033***	0.038***	0.038***
	(0.007)	(0.007)	(0.006)	(0.006)
Mothers' economic and social independence				
Manage own money	0.029***	0.026***	0. 022**	0.024***
	(0.008)	(0.009)	(0.011)	(0.009)
Participates in voluntary activities	0.008	0.013	0.005	0.005
	(0.008)	(0.009)	(0.012)	(0.009)
Currently studying	0.009*	0.010*	0.013***	0.013***
	(0.005)	(0.005)	(0.005)	(0.005)
Own or joint decision on own work status	0.014***	0.017***	0.020***	0.022***
	(0.004)	(0.005)	(0.008)	(0.007)
Household decisions- making				
Own/joint decision on child's education	0.011***	0.011**	0.015**	0.016***
	(0.004)	(0.004)	(0.006)	(0.005)
Own/joint decision on own health	0.002	0.002	-0.001	-0.000
	(0.005)	(0.006)	(0.006)	(0.006)
Own/joint decision on child's discipline	0.012**	0.013**	0.019**	0.018**
	(0.005)	(0.006)	(0.009)	(0.008)
Own/joint decision on expenditure	0.018**	0.018**	0.020	0.021**
	(0.009)	(0.009)	(0.013)	(0.010)
Own/joint decision on food expenditure	0.009	0.009	0.008	0.010
	(0.008)	(0.009)	(0.012)	(0.010)
Own/joint decision on important matters	0.009	0.010	0.002	0.005
	(0.008)	(0.009)	(0.009)	(0.008)
Own/joint decision on having children	0.003	0.004	-0.001	0.000
	(0.004)	(0.004)	(0.003)	(0.004)
Own/joint decision on contraceptives	-0.004	-0.001	-0.002	-0.002
	(0.005)	(0.006)	(0.007)	(0.006)
Own/joint decision on own health	-0.002	-0.003	-0.009	-0.009
	(0.006)	(0.007)	(0.005)	(0.005)
Own/joint decision on if mothers can visit	-0.002	-0.000	0.002	0.003
	(0.008)	(0.008)	(0.011)	(0.009)
Child Controls	No	Yes	No	Yes
Household Demographics	No	Yes	No	Yes
Household Economics	No	Yes	No	Yes
Observations	496	496	496	496

Note. Each cell reports the estimated effect of years of treatment on the mothers' outcome from a separate regression based on the 2012 and 2013 survey. Columns (1)-(2) present results using the original sample without entropy balancing. Columns (3)-(4) stem from the weighted sample adjusted by entropy balancing. Standard errors are presented in parentheses and are clustered at the maternal level; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C.10: ESTIMATED EFFECTS ON MOTHERS' OUTCOMES: SUMMARY INDICES (BASED ON THE SAMPLE OF MOTHERS ENROLLED IN THE PROGRAM BEFORE 2007)

	Not weighted			Weighted		
	(1)	(2)	(3)	(4)	(5)	(6)
A: Sample based on 2012 survey						
Labor market outcomes	0.564***	0.533***	0.540***	0.629***	0.618***	0.619***
	(0.110)	(0.108)	(0.106)	(0.121)	(0.112)	(0.104)
Economic and social independence	0.323***	0.327***	0.360***	0.297***	0.309***	0.341***
	(0.075)	(0.076)	(0.076)	(0.082)	(0.075)	(0.068)
Intra-household decision-making	0.180***	0.177***	0.195***	0.192***	0.191***	0.194***
	(0.060)	(0.061)	(0.061)	(0.068)	(0.060)	(0.057)
Child's investment	0.155**	0.194**	0.193**	0.197**	0.211**	0.198**
	(0.071)	(0.073)	(0.075)	(0.092)	(0.088)	(0.085)
Self-esteem and Big Five Personality Traits	0.015	0.037	0.040	0.033	0.035	0.036
	(0.076)	(0.078)	(0.078)	(0.092)	(0.085)	(0.081)
Observations	224	224	224	224	224	224
B: Pooled sample of 2012 and 2013						
Labor market outcomes	0.560***	0.524***	0.514***	0.605***	0.602***	0.598***
	(0.112)	(0.119)	(0.110)	(0.125)	(0.115)	(0.111)
Economic and social independence	0.228***	0.231***	0.248***	0.220***	0.231***	0.236***
	(0.058)	(0.056)	(0.058)	(0.066)	(0.061)	(0.059)
Intra-household Decision-making	0.136**	0.149***	0.150***	0.124*	0.123**	0.123**
	(0.055)	(0.054)	(0.055)	(0.064)	(0.055)	(0.052)
Observations	394	394	394	394	394	394
Child Controls	No	Yes	Yes	No	Yes	Yes
Household Demographics	No	Yes	Yes	No	Yes	Yes
Household Economics	No	No	Yes	No	No	Yes

NOTE. Each cell reports the estimated mean effect from a separate regression. Column (1)-(3) present results using the original sample without entropy balancing. Column (4)-(6) stem from the weighted sample adjusted by entropy balancing. Standard errors are presented in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C.11: ESTIMATED EFFECTS ON MOTHERS' OUTCOMES: SUMMARY INDICES (BASED ON THE SAMPLE OF MOTHERS ENROLLED IN THE PROGRAM FROM 2007)

	Not weighted			Weighted		
	(1)	(2)	(3)	(4)	(5)	(6)
A: Sample based on 2012 survey						
Labor market outcomes	0.399***	0.354***	0.413***	0.677***	0.668***	0.687***
	(0.121)	(0.130)	(0.125)	(0.186)	(0.156)	(0.141)
Economic and social independence	0.205**	0.230**	0.250**	0.147	0.136	0.236
	(0.101)	(0.111)	(0.111)	(0.138)	(0.155)	(0.146)
Intra-household decision-making	0.039	-0.042	-0.025	-0.227	-0.274*	-0.274*
	(0.080)	(0.090)	(0.090)	(0.189)	(0.148)	(0.144)
Child's investment	0.227**	0.232**	0.219**	0.235**	0.215**	0.191**
	(0.091)	(0.106)	(0.108)	(0.115)	(0.100)	(0.085)
Self-esteem and Big Five Personality Traits	0.071	0.061	0.056	0.259	0.276**	0.283***
	(0.095)	(0.106)	(0.106)	(0.221)	(0.111)	(0.104)
Observations	168	168	168	168	168	168
B: Pooled sample of 2012 and 2013						
Labor market outcomes	0.549***	0.453***	0.493***	1.035***	1.037***	1.028***
	(0.138)	(0.149)	(0.140)	(0.206)	(0.171)	(0.155)
Economic and social independence	0.227***	0.296***	0.297***	0.536***	0.571***	0.582***
	(0.083)	(0.090)	(0.091)	(0.194)	(0.175)	(0.163)
Intra-household Decision-making	0.058	-0.023	-0.029	-0.273	-0.340*	-0.345**
	(0.074)	(0.077)	(0.078)	(0.214)	(0.174)	(0.164)
Observations	292	292	292	292	292	292
Child Controls	No	Yes	Yes	No	Yes	Yes
Household Demographics	No	Yes	Yes	No	Yes	Yes
Household Economics	No	No	Yes	No	No	Yes

NOTE. Each cell reports the estimated mean effect from a separate regression. Column (1)-(3) present results using the original sample without entropy balancing. Column (4)-(6) stem from the weighted sample adjusted by entropy balancing. Standard errors are presented in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C.12: Estimated Effects on Children's Outcomes: Summary Indices (2012) (Based on the Treatment Sample Enrolled Before 2007)

	Not weighted			Weighted		
	(1)	(2)	(3)	(4)	(5)	(6)
Overall summary index†	0.155**	0.130**	0.138**	0.128	0.140*	0.143*
	(0.059)	(0.065)	(0.063)	(0.080)	(0.080)	(0.080)
Tests scores	0.129	0.142	0.151	0.014	0.146	0.144
	(0.109)	(0.107)	(0.110)	(0.153)	(0.115)	(0.115)
Schooling dropout and grade repetition	–0.170*	–0.150*	–0.143*	–0.166	0.187*	–0.187*
	(0.087)	(0.082)	(0.077)	(0.111)	(0.112)	(0.110)
Attitude towards schooling	0.085	0.021	0.019	0.102	0.017	0.013
	(0.066)	(0.076)	(0.075)	(0.073)	(0.081)	(0.078)
Child Controls	No	Yes	Yes	No	Yes	Yes
Household Demographics	No	Yes	Yes	No	Yes	Yes
Household Economics	No	No	Yes	No	No	Yes
Observations	313	313	313	313	313	313

Note. Each cell reports the estimated effect on a summary index from a separate regression based on the 2012 survey. Column (1)-(3) present results using the original sample without entropy balancing. Column (4)-(6) stem from the weighted sample adjusted by entropy balancing. Standard errors are presented in parentheses and are clustered at the maternal level; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. † Signs of outcomes of schooling dropout and grade repetition are reversed when calculating the overall summary index.

Table C.13: ESTIMATED EFFECTS ON CHILDREN'S OUTCOMES: SUMMARY INDICES (2012) (BASED ON THE TREATMENT SAMPLE ENROLLED FROM 2007)

	Not weighted			Weighted		
	(1)	(2)	(3)	(4)	(5)	(6)
Overall summary index†	0.166**	0.119	0.109	0.090	0.131	0.201
	(0.082)	(0.087)	(0.094)	(0.095)	(0.109)	(0.129)
Tests scores	0.141	0.150	0.087	0.264	0.229	0.026
	(0.150)	(0.142)	(0.137)	(0.193)	(0.154)	(0.130)
Schooling dropout and grade repetition	–0.265***	–0.116	–0.145	–0.058	–0.100	–0.131
	(0.077)	(0.078)	(0.094)	(0.037)	(0.075)	(0.087)
Attitude towards schooling	0.097	0.066	0.066	0.095	-0.030	-0.085
	(0.096)	(0.110)	(0.115)	(0.114)	(0.147)	(0.146)
Child Controls	No	Yes	Yes	No	Yes	Yes
Household Demographics	No	Yes	Yes	No	Yes	Yes
Household Economics	No	No	Yes	No	No	Yes
Observations	229	229	229	229	229	229

NOTE. Each cell reports the estimated effect on a summary index from a separate regression based on the 2012 survey. Column (1)-(3) present results using the original sample without entropy balancing. Column (4)-(6) stem from the weighted sample adjusted by entropy balancing. Standard errors are presented in parentheses and are clustered at the maternal level; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

† Signs of outcomes of schooling dropout and grade repetition are reversed when calculating the overall summary index.