UNIVERSITÀ POLITECNICA DELLE MARCHE

DIPARTIMENTO DI SCIENZE ECONOMICHE E SOCIALI

BAYESIAN MODEL AVERAGING FOR PROPENSITY SCORE MATCHING IN TAX REBATE

RICCARDO (JACK) LUCCHETTI LUCA PEDINI
CLAUDIA PIGINI

QUADERNO DI RICERCA n. 457
ISSN: 2279-9575

June 2021

Collana curata da:
Massimo Tamberi

Abstract

Propensity Score Matching is a popular approach to evaluate treatment effects in observational studies. However, when building the underlying propensity score model practitioners often overlook the issue of model uncertainty and its consequences. We tackle this problem by Bayesian Model Averaging (BMA) with an application to the 2014 Italian tax credit reform (the so-called "Renzi bonus"). Model uncertainty has a great impact on the estimated treatment effects. BMA-based estimates point towards a significant effect of the rebate on food consumption only for liquidity constrained households; conversely, model selection procedures sometimes produce results incompatible with the consumption smoothing hypothesis.

JEL Class.: C11, C52, D12

Keywords: 2014 Italian Tax Credit Reform, Bayesian Model Averaging, Model Uncertainty, Propensity Score Matching, Reversible Jump Markov Chain Monte Carlo, Tax Rebate Policies

Indirizzo: Riccardo (Jack) Lucchetti: Department of Economics and Social Sciences, Marche Polytechnic University, Ancona (IT). E-mail: r.lucchetti@staff.univpm.it
Luca Pedini: Department of Economics and Social Sciences, Marche Polytechnic University, Ancona (IT). E-mail: l.pedini@pm.univpm.it
Claudia Pigini: Department of Economics and Social Sciences, Marche Polytechnic University, Ancona (IT). E-mail: c.pigini@staff.univpm.it

Bayesian Model Averaging for Propensity Score Matching in Tax Rebate

Riccardo (Jack) Lucchetti *Luca Pedini* *Claudia Pigini*

1 Introduction

Propensity Score Matching (PSM) (Rosenbaum and Rubin, 1983) is a standard tool for treatment evaluation in observational studies and it has been used in a wide range of economic applications. However, the specification of the binary choice model used to compute the Propensity Score (PS), namely the conditional probability of being treated given a chosen set of characteristics, is often built with little attention to the choice of explanatory variables, despite the fact that the selection of controls can dramatically affect the matching and, as a result, the estimate of the causal effect (Heckman et al., 1997; Caliendo and Kopeinig, 2008).

From a theoretical standpoint, the consequences of model selection can be assessed in terms of the so-called *bias-variance trade-off*. Specifying the linear index using all the available covariates may be tempting so as to gain a greater model flexibility and reduce the approximation bias. However, over-parametrisation may lead to violations of basic assumptions of PSM, such as the common support and balancing conditions, as well as to overfitting and to a larger variance of the estimator. Parsimonious models, on the other hand, may produce an inconsistent treatment effect estimator because of relevant omitted variables (Heckman et al., 1997).

Economic theory may provide some background information to guide the practitioner in the choice of covariates, albeit on a case-wise basis and depending on the specific treatment analysed. More general strategies have been put forward by the related stream of literature to help overcome this impasse. Rosenbaum and Rubin (1984) suggest using the PS along with stratification matching and adjusting the set of covariates with interactions or higher order polynomials, as long as the balance conditions are met. The same recommendations are made by Dehejia and Wahba (1999). Rubin and Thomas (1996) points out that discarding covariates should be always a mindful decision motivated by the balancing conditions. Millimet and Tchernis (2009) emphasise instead the advantage of estimating redundant models over parsimonious ones whenever the bias reduction is larger than the increase in variance, arguing that overfitting has no negative consequences.

Bryson et al. (2002) warn on the opposite problem of including too many variables, as efficiency would be often affected in empirical applications. In addition, they also question the quality of the resultant matching when redundant variables are included.

There is also a related stream of literature dealing with treatment effects evaluation in cases where the number of covariates is larger than the number of observations and where model selection is tackled via the *sparsity* assumption and LASSO-like estimators (Belloni et al., 2014, 2017; Chernozhukov et al., 2018). Finally, Brookhart et al. (2006) take an intermediate position between parsimonious and inclusive specifications: variables that are unrelated to the treatment assignment but related to the outcome should always be included, whereas the converse is not necessarily true.

The lack of a unique modelling strategy for the PS is a consequence of the impossibility of choosing the *true* model specification, which clearly makes any kind of model selection procedure, aimed to achieve an ideal bias-variance trade-off, inadequate. As a consequence, a problem of *model uncertainty* (Chatfield, 1995) is introduced by conditioning the analysis on a single model specification, under the assumption that such specification is the true one.

Alternatively, model uncertainty can be dealt with by averaging estimates from a range of different models, using weights that reflect the models' goodness of fit. Model averaging can be performed either in a frequentist or a Bayesian framework, where the latter provides an immediate interpretation of weights as model probabilities and a direct measure of the relative importance of covariates in the model specification. Moreover, Bayesian Model Averaging (BMA) offers a computational advantage in high dimensional model spaces as it is often based on Markov Chain Monte Carlo (MCMC) samplers. Finally, the choice of the prior distributions can be exploited to perform model regularisation. See Steel (2020) for a detailed discussion on the advantages of BMA over the frequentist model averaging approach.

Model averaging for PSM has been considered by Kitagawa and Muris (2016) and Xie et al. (2019) within the frequentist framework, whereas other solutions pertain specifically to BMA. In particular, Kaplan and Chen (2014) propose a two-step strategy in which PS model probabilities are first approximated by the Bayesian Information Criterion (BIC) and used to sample model specific parameters by MCMC, that are then employed to compute the PS and the resulting treatment effect. Notice that this approach is not based on simultaneous sampling of models and parameters, which results in computational inefficiency and into low estimation accuracy, also due to the use of BIC approximation.

Zigler and Dominici (2014) overcome these limitations by means of the Stochastic Search Variable Selection (SSVS) (George and McCulloch, 1993), which allows for the simultaneous sampling of model and parameters. PSM is framed within the control-function approach: the outcome equation is specified as a function of both the treatment and the PS, along with a set of controls. However, this gives rise to the so-called "feedback effect": since some covariates may appear in both the equations, for the PS and outcome, posterior distributions for the outcome models can affect the PS ones. This issue has been extensively debated, as it can severely undermine PS properties and reliability (see also Cefalu et al., 2017).

In this paper, we propose to address the issue of model uncertainty in PSM by a BMA approach that improves upon Kaplan and Chen (2014) and Zigler and Dominici (2014) as it avoids feedback effects and maintains a high degree of computational efficiency and accuracy. To this end, we employ the Reversible Jump MCMC (RJMCMC) sampler (Green, 1995; Holmes and Held, 2006; Lamnisos et al., 2009), which makes it possible to simultaneously sample the models and their parameters, and consequently the PS distribution across the model space; no two-step procedures are needed and, at the same time, BIC approximations are avoided. Moreover, we consider PSM as a three-step procedure, where the three steps are (a) the PS model (b) matching and (c) the treatment effect evaluation. In this way, the model uncertainty issue directly affects only the PS and the related model probabilities can be attached to each treatment effect estimator.

Our proposed method is used for evaluating the economic impact of the 2014 Italian tax credit reform (Decree Law 66/2014), also known as "Renzi bonus", which introduced a monthly wage increase of about €80 for all employees with an annual gross income between €8145 and €26000. Tax reductions are often employed as means to counteract the negative effects of recessions, as an increase in disposable income would supposedly encourage household consumption. Yet the effectiveness of tax credit policies is widely debated (Shapiro and Slemrod, 2003a,b, 2009). Furthermore, empirical evidence from micro data has been found to be rather sensitive to methodological choices (Heim, 2007).

We use data from the Survey of Household Income and Wealth (SHIW) issued by the Bank of Italy and estimate the treatment effect of the tax rebate on different types of monthly consumption, such as total, food, non durables, car-related, and other durables consumption. In line with the results provided by Neri et al. (2017), who evaluated the effect of the Renzi bonus using the same data, we find a role of liquidity constraints, as the treatment effect is significant only on the food consumption of households who report greatest difficulty to make ends meet. In addition, we show that accounting for model uncertainty leads to different evidence than that produced by a standard model selection procedure, which would point toward an increase in food and other non durables consumption for all households.

The rest of the paper is organised as follows: in Section 2 we briefly introduce the mechanisms driving the consumption effects of tax rebate and illustrate the Italian 2014 policy intervention; Section 3 describes the proposed BMA for PSM; Section 4 introduces the SHIW dataset and reports the relevant descriptive statistics; Section 5 provides the estimation results; Section 6 discusses the results and contains the final remarks.

2 Background

Tax rebate policies are common stabilising instruments applied by policymakers to reduce the impact of the business cycle: these fiscal interventions are aimed to increase household spending, thus countering the negative effect of recessions. Their effects on consumption, however, might be difficult to measure. On the one hand, the permanent income hypothesis predicts that transitory policy interventions should not affect consumption: if individuals smooth consumption over their lifetime, a tax rate is effective only as long as is perceived as permanent. Moreover, the intervention might be anticipated, which makes it more difficult to analyse if consumers adjust their behaviour before actual implementation. On the other hand, if households face liquidity constraints, an increase in disposable income could raise consumption towards its life-cycle/desired level, thereby generating a positive effect. For a review of the related theoretical literature, see Jappelli and Pistaferri (2010).

Empirical literature on the effects of tax reductions on household consumption has relied either on dedicated surveys, where respondents are asked how they wish to use their extra funds (see Neri et al., 2017, for a review) or on expenditure survey microdata, such as the one employed in this paper. As for the evaluations on general household consumption and expenditure surveys, Wilcox (1989), Parker (1999), and Souleles (2002) find a positive effect of tax cuts on consumption using a difference-in-difference approach. The findings provided by Souleles (1999) suggest that tax refunds are used for non-durables by low-income/low-liquidity households, thus finding evidence of the role of liquidity constraints, whereas high-income/high-liquidity individuals seem to prefer durables expenditure. Johnson et al. (2006) find confirmation of a positive effect of tax rebate policies, whose magnitude seems to be more pronounced in the case of liquidity constraints. According Agarwal et al. (2007), consumers initially tend to save the bonus, by increasing their credit line and thereby paying off debt, while an increase in spending occurs afterwards. Finally, Parker et al. (2013) find that households spend most of the rebate in durables.

Heim (2007) warns that the insights from these empirical findings should be interpreted with care: results tend to vary depending on the selected sample, the targeted expenditure component, and the set of regressors used in the treatment evaluation model. It is worth recalling that the positive effect on consumption of tax interventions is only partly supported by the theoretical background and its evidence in empirical studies might well be the result of non-obvious modelling choices. Conclusions on the effectiveness of these policies should therefore drawn with caution, considering how costly they may be.

According to Government estimates, the 2014 Italian Decree Law 66/2014, also known as "Renzi bonus", entailed a total transfer of €5.9 billion to households (for about 10 million employees), which amounts to 0.4% of GDP. From a technical point of view, the tax rebate was translated into an increase in the monthly salary or pension by €80. Recipients were payroll employees or retired workers with a total annual income between €8145 and €26000. The rebate was delivered directly in the monthly check, starting in May 2014. Because the eligibility status was based on the 2014 income, which has been verified only in 2015, several cases of classifications of people near the lower threshold have occurred. In fact, it was estimated that about 1.5 million people were incorrectly classified as eligible, and had to reimburse their bonus. Finally, it is worth noting that the standard bonus amount of €80 was reduced for those people whose income was €24000-26000 and for workers who found a job in 2014, as the bonus was proportional to the number of months spent in employment during the year.

Neri et al. (2017) provide some empirical evidence of the 2014 tax rebate based on the SHIW dataset, by extracting a panel for the years 2012 and 2014, using PSM to match treated and control groups and finally evaluating the treatment effect by difference-in-difference. On the one hand, this way of performing difference-in-difference eliminates some between-group heterogeneity, thus making the parallel trend assumption likely to be satisfied in practice. On the other, the procedure can be viewed as pure PSM where the outcome is the difference between consumption/expenditure in 2014 and in 2012. The estimated change in household expenditure amounts to 50%–60% of the sum received, with larger responses for low-liquidity/low-income households, thus confirming a role for liquidity constraints.

3 Methods

3.1 Propensity Score Matching and Model Uncertainty

Let D_i be a treatment binary indicator for the i-th individual, $i = 1, \ldots, n$, with $D_i = 1$ for treated subjects. Covariates with pre-treatment individual and household information are collected into the $n \times k$ matrix X. The number of possible subsets of the columns of X is 2^k; let $\mathcal{M}$ be set of models (the model space), with M_j being its j-th member ($j = 1, \ldots, 2^k$); also, define X_j be the matrix containing the corresponding subset of columns of X.

The PS, that is the conditional probability of being treated given a chosen set of characteristics, is usually described via a generalised linear model of the form

$$g\left(\mathrm{E}\left[D_i|\boldsymbol{x}_{ij}\right]\right) = \boldsymbol{x}'_{ij}\boldsymbol{\beta}_j, \tag{1}$$

where $g(\cdot)$ is an appropriate link function associated with a binary choice model, $\boldsymbol{x}_{ij}$ is the vector of covariates for individual i from matrix X_j, and $\boldsymbol{\beta}_j$ is the vector of parameters specific to M_j.

PSM is traditionally performed sequentially, where first the PS is computed via Maximum-Likelihood (ML) estimation of $\boldsymbol{\beta}_j$ for a chosen M_j, then matching of treated with untreated individuals is performed on the basis of the estimated PS and, finally, the treatment effect of interest is evaluated as the mean difference in the outcome y between matched units. In particular, in the following we consider the average treatment effect on the treated (ATET), which is often considered the most relevant quantity in empirical applications and measures the effects on individuals for whom the treatment is intended. Standard unconfoundedness and overlap conditions, needed for the identification of the ATET, are assumed to hold hereafter.

Following the notation in Caliendo and Kopeinig (2008), the PSM estimator of the ATET can be viewed as the estimator of

$$\gamma_j = \mathrm{E}_{\mathrm{Pr}(\boldsymbol{x}_{ij})|D_i=1}\{\mathrm{E}\left[y_i(1)|D_i = 1, \mathrm{Pr}(\boldsymbol{x}_{ij})\right] - \mathrm{E}\left[y_i(0)|D_i = 0, \mathrm{Pr}(\boldsymbol{x}_{ij})\right]\}, \tag{2}$$

where $\mathrm{Pr}(\boldsymbol{x}_{ij}) = \mathrm{E}\left[D_i|\boldsymbol{x}_{ij}\right]$ is the PS and $y_i(1)$ and $y_i(0)$ are the so-called potential outcomes. Since the ATET depends on the initial choice of the PS model M_j, the γ parameter also bears the subscript j.

As discussed in Section 1, the choice of using one model M_j implies an assumption on a *true* model specification, to which the parameter γ_j corresponds. If model uncertainty is taken into account, the possibility to choose among the 2^k specifications in $\mathcal{M}$ calls for a model averaging procedure. By exploiting the sequential nature of PSM, it is possible to link directly the uncertainty affecting the PS model to the subsequent treatment effect estimation. In a frequentist setting, a model averaging estimator $\hat{\gamma}_{MA}$ is given by averaging model specific treatment effect estimates $\hat{\gamma}_j$ across the model space $\mathcal{M}$. Formally, we have

$$\hat{\gamma}_{MA} = \sum_{j=1}^{2^k} \hat{\gamma}_j \omega_j, \tag{3}$$

where ω_j is the model's weight.

Equation (3) is translated into the BMA framework by considering M_j and γ_j as random variables: each model-specific weight is replaced by the posterior probability for the PS model, which can be expressed as

$$\mathrm{P}(M_j|\mathcal{D}) = \frac{\mathrm{p}(\mathcal{D}|M_j)\mathrm{P}(M_j)}{\mathrm{P}(\mathcal{D})},$$

where $\mathcal{D}$ is the data (treatment assignment, outcome variable, covariates), $\mathrm{p}(\mathcal{D}|M_j)$ is the marginal data density (also known as marginal likelihood) and $\mathrm{P}(M_j)$ is the prior distribution for model M_j. Following Zigler and Dominici (2014), who define the model averaging posterior distribution for the causal effect, the treatment evaluation problem for the PSM ATET parameter γ can be expressed in terms of posterior expected values of the model-specific γ_j as

$$\mathrm{E}(\gamma|\mathcal{D}) = \sum_{j=1}^{2^k} \mathrm{E}(\gamma_j|\mathcal{D}, M_j)\mathrm{P}(M_j|\mathcal{D}).$$

By exploiting the "intermediate Bayesian" methodology in An (2010), further discussed by Kaplan and Chen (2012), we can obtain the estimated counterpart of the above equation as

$$\hat{\gamma}_{BMA} = \sum_{j=1}^{2^k} \hat{\gamma}_j \hat{\mathrm{P}}(M_j|\mathcal{D}), \tag{4}$$

where $\hat{\gamma}_j$ is a frequentist estimator of the ATET in (2) and $\hat{\mathrm{P}}(M_j|\mathcal{D})$ is the estimated posterior model probability.

In practice, Equation (4) is problematic to compute. First, it is analytically intractable because in general there is no closed formula for model weights. Second, for realistic values of k the cardinality of the model space is too large to make direct computation feasible. Finally, posterior probabilities must be estimated not only for the PS models $\mathrm{P}(M_j|\mathcal{D})$ but also for the parameters $\boldsymbol{\beta}_j$ in (1) for which closed-form posterior probabilities are generally not available either.

The solutions put forward in the literature range from simple approximations (such as using the BIC for the model posterior probabilities and the ML estimate as the mean of parameters posteriors), to more sophisticated approaches such as the conjugate prior method (Chen et al., 2008), the SSVS (George and McCulloch, 1993), and the RJMCMC (Green, 1995), which is what we adopt here.

In standard BMA applications (Madigan et al., 1995; Hoeting et al., 1999) the model and related parameters can only be retrieved via two-step procedures: model posteriors are computed first, and in a second step the related parameters are obtained via an additional MCMC from each sampled model. In the RJMCMC approach, instead, the model and related parameters are sampled simultaneously at each MCMC iteration and the model averaging posterior distribution of the parameter vector is immediately available. Furthermore, approximations of the model posteriors are avoided and the exploration of the full model space reduces to those models with a higher posterior probability.

The RJMCMC approach overcomes many limitations of other approaches. Compared to Kaplan and Chen (2014), we avoid BIC approximations of model posteriors as well as two-step procedures. Furthermore, the standard sequential framework for PSM estimation shields our procedure from feedback effects, that may hamper estimation of the ATET in the Zigler and Dominici (2014) approach. Moreover, since parameter posteriors are available, the sample of parameters from RJMCMC can be used to compute a sample of PSs, whose distribution encompass the variability within each single model and between different models. Finally, compared to the SSVS approach, RJMCMC does not require strong parametric assumptions on the the priors and link function specification. This makes SSVS computationally more convenient; in this application, however, we mitigate the greater computational burden of RJMCMC by CPU parallelisation.

3.2 The RJMCMC technique

Formally, the RJMCMC technique can be seen as a variation of the standard Metropolis-Hastings scheme (Hastings, 1970), in which the vector of parameters of interest $\boldsymbol{\beta}$ is allowed to change its dimension across simulations. At each MCMC iteration h, a model $M^{(h)}$ is proposed from the previous sampled $M^{(h-1)}$ by means of a proposal distribution $q(M^{(h)}|M^{(h-1)})$. The corresponding parameter $\boldsymbol{\beta}^{(h)}$ is derived by simply transforming the ones of the previous model $\boldsymbol{\beta}^{(h-1)}$ via a differential function $f(\cdot)$, so that $\boldsymbol{\beta}^{(h)} = f(\boldsymbol{\beta}^{(h-1)})$. Since $\boldsymbol{\beta}^{(h)}$ and $\boldsymbol{\beta}^{(h-1)}$ may be vectors of different size, a matching variable u is used. Suppose for example to have $k = 3$ covariates $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$ with the related set of parameters $\beta_1, \beta_2, \beta_3$ and assume to start, at iteration $h-1$, from $M^{(h-1)} = (\boldsymbol{x}_1, \boldsymbol{x}_2)$. If the proposal distribution draws a new specification by adding $\boldsymbol{x}_3$, then $M^{(h)} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)$. Since the dimension of the new model is larger, in order to compute $\boldsymbol{\beta}^{(h)} = (\beta_1^{(h)}, \beta_2^{(h)}, \beta_3^{(h)})$ an additional term u is needed: $\boldsymbol{\beta}^{(h-1)} = (\beta_1^{(h-1)}, \beta_2^{(h-1)}) \rightarrow \boldsymbol{\beta}^{(h)} = f(\beta_1^{(h-1)}, \beta_2^{(h-1)}, u^{(h)})$.

The probability of accepting a movement from $(M^{(h-1)}, \boldsymbol{\beta}^{(h-1)})$ to $(M^{(h)}, \boldsymbol{\beta}^{h})$ is

$$\alpha = \min\left[1, \frac{\mathrm{P}(M^{(h)}, \boldsymbol{\beta}^{(h)}|\mathcal{D})q(M^{(h-1)}|M^{(h)})}{\mathrm{P}(M^{(h-1)}, \boldsymbol{\beta}^{(h-1)}|\mathcal{D})q(M^{(h)}|M^{(h-1)})}\left|\frac{\partial f(\boldsymbol{\beta}^{(h)}, M^{(h)}, u^{(h)})}{\partial(\boldsymbol{\beta}^{(h)}, M^{(h)}, u^{(h)})}\right| z(u^{(h)})\right]$$

where $\mathrm{P}(M^{(h)}, \boldsymbol{\beta}^{(h)}|\mathcal{D})$ is the target distribution, i.e. the joint posterior for the couple $(M^{(h)}, \boldsymbol{\beta}(h))$, $\left|\frac{\partial f(\boldsymbol{\beta}^{(h)}, M^{(h)}, u^{(h)})}{\partial(\boldsymbol{\beta}^{(h)}, M^{(h)}, u^{(h)})}\right|$ is the Jacobian term, and $z(u^{(h)})$ is given by the ratio of the density distributions of u referring respectively to $M^{(h)}$ and $M^{(h-1)}$. In this paper, we use the implementation provided in the `gretl` package `ParMA`; we refer the reader to Lucchetti and Pedini (2020) for an illustration of the package as well as further details on RJMCMC and its application to generalised linear models.

For the PSM estimation procedure at hand, the issue of model uncertainty for the PS model is handled by using the fact that RJMCMC delivers a sample of $(\boldsymbol{\beta}^{(h)}, M^{(h)})$, $h = 1, \ldots, H$. Each parameter is used to compute a PS and the related treatment effect $\hat{\gamma}^{(h)}$. To obtain the counterpart of Equation (4), we simply compute the sample mean as

$$\hat{\gamma}_{BMA} = \frac{\sum_{h=1}^{H} \hat{\gamma}^{(h)}}{H}. \tag{5}$$

For the posterior variance $V(\gamma|\mathcal{D})$, we can use

$$\hat{V}(\gamma|\mathcal{D}) = \frac{\sum_{h=1}^{H} \hat{\sigma}^{2^{(h)}}}{H} + \frac{\sum_{h=1}^{H}(\hat{\gamma}^{(h)} - \hat{\gamma}_{BMA})^2}{H-1}, \tag{6}$$

where the first element on the right hand-side is the sample mean of the variances $\hat{\sigma}^{2^{(h)}}$ of the treatment effects, and the second one is the sample variance of the treatment estimates. In this way, we account not only for the variability of $\hat{\gamma}$ due to the model uncertainty and the PS, but also for the inner variability of the parameter γ itself.

4 The data

For our empirical analysis, we use Bank of Italy's SHIW dataset. The SHIW collects micro-data on the economic and financial behaviour of Italian households with biannual periodicity. It covers a sample of individuals belonging to about 8000 households across 300 municipalities, about half of which are panel households who were also interviewed in earlier waves. Households are selected randomly in two stages: first, municipalities are chosen from a stratified sample by region and population guaranteeing that panel units are included, then households are drawn from each municipality. Two datasets are available, that can be matched at the individual and household level. One dataset is cross-sectional, which includes additional information specific to certain waves, such as special modules for research purposes or inquiries about the effects of new reforms; the other one is a historical dataset, that includes harmonised waves for a subset of households since 1989.

Table 1: Outcome variables description

Variable	Description
c_tot	Total consumption (monthly)
c_food	Food consumption (monthly)
c_notd	Total consumption of non durable goods (monthly)
c_car	Total consumption of cars/means of transportation (monthly)
c_otherd	Total consumption of other durable goods (monthly)

Here, we use the 2014 wave and the historical panel data for the years 2012 and 2014. The cross-section dataset includes information about the 2014 tax credit: whether the interviewed household received the bonus, how many components were benefited, the amount received and spent. As pointed out by Neri et al. (2017), this particular part of the survey mirrors the one in Shapiro and Slemrod (2003a), who introduced the use of survey responses to analyse tax rebate impacts. The panel datasets allows us instead to access information on both pre- and post-treatment information. For a group of panel households (some of which eligible for the bonus) the consumption pattern can be observed across the two years and the impact of the bonus can thus be inferred. By joining the two data sources, we build a sample of $4,459$ households observed in 2012 and 2014, of which 864 were eligible for the bonus (19.4%). Individual observations refer to households, so the assignment status indicates whether at least one person is eligible.

The outcome variables we consider are: total consumption, food consumption, non-durables expenditure, car-related consumption, and other durables (furniture, appliances). All variables are monthly. The definitions of the outcome variables are listed in Table 1 and Table 2 reports the related summary statistics for the selected sample. Total consumption decreased from 2012 to 2014 in both treated and control units. As for specific categories, it is interesting to note that food consumption, car consumption and other durables increased in 2014 for the treated group, whereas non durable consumption decreased.

The covariates used to build the PS model are described in Table 3 and their summary statistics are reported in Table 4. These variables are relative to the pre-treatment period 2012 and reflect individual socio-cultural and economic condition.

As it can be seen from Table 4, most tax credit recipients belong to the blue-collar and office worker category. However, the high share of "not currently employed" individuals (either unemployed, or retired, or otherwise outside the labour force) in the treated group deserves some clarification. An explanation comes from the definition of QUALP10: the variable refers only to the declared head of household, which may differ from the actual recipient. From the 2014 wave, it is impossible to discern which component inside the household is the true receiver since the binary treatment assignment refers to the household as a whole.

Table 2: Outcome variables: main summary statistics

	2014			2012		
	Full sample	treated	control	Full sample	treated	control
c_tot						
mean	1948.3	2129.0	1904.9	2149.2	2326.7	2106.5
std.dev.	1174.3	1031.3	1202.3	1276.3	1107.8	1310.2
median	1683.3	1875.0	1600.0	1858.3	2107.5	1791.7
min	133.33	300.00	133.33	-583.33	-416.67	-583.33
max	15617	10742	15617	15867	11937	15867
c_food						
mean	545.08	616.61	528.37	544.95	597.85	532.23
std.dev.	295.72	281.05	296.75	301.64	283.55	304.50
median	500.00	600.00	500.00	500.00	550.00	500.00
min	50.000	100.00	50.000	50.000	50.000	50.000
max	3500.0	2500.0	3500.0	3000.0	2700.0	3000.0
c_notd						
mean	1866.1	1989.8	1836.4	2065.8	2224.4	2027.7
std.dev.	1056.1	853.25	1097.3	1175.7	1021.1	1206.9
median	1600.0	1800.0	1600.0	1800.0	2045.0	1745.0
min	150.00	550.00	150.00	150.00	500.00	150.00
max	14200	6750.0	14200	15867	11012	15867
c_car						
mean	51.473	94.736	41.076	49.378	58.688	47.141
std.dev.	283.72	392.48	249.60	274.60	271.86	275.24
median	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
min	-1000	-666.67	-1000	-2416.7	-2416.7	-1583.3
max	5166.7	5166.7	4166.7	4333.3	2083.3	4333.3
c_otherd						
mean	30.745	44.421	27.458	34.031	43.662	31.717
std.dev.	141.04	159.27	136.11	117.06	136.67	111.73
median	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
min	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
max	4166.7	2500.0	4166.7	2750.0	1666.7	2750.0

Note: Negative value in car consumption are due the definition of the variable, which includes depreciation.

Table 3: Covariates description

Variable	Description
NEQU	number of components in the household (normalised by age)
CLY2	categorical variable for income class (5 categories by quintiles)
AREA3	categorical variable for geographical area (3 categories)
ETA5	categorical variable for head of household age class (5 categories)
STUDIO	categorical variable for head of household educational level (6 categories)
QUALP	categorical variable for head of household employment activity (10 categories)
CONDGEN	categorical variable for perceived economic hardship (6 categories)

Bonus receivers seem to hold mainly to middle school - high school level of educational attainment, while the income class is rather heterogeneous. As for the geographical area, about 50% of the units live in the North. The CONDGEN variable is particularly important here: it contains an indicator of perceived economic hardship. Treated units belong mostly to the middle category (moderate or little economic problems).

5 Estimation results

5.1 PS models

For the RJMCMC application presented here we use a Probit model for the PS, with a uniform prior distribution for models, $\mathrm{P}(M_j)$, and a Zellner-g type (Zellner and Siow, 1980) for the PS model parameter vector $\boldsymbol{\beta}_j$.

These prior choices are standard in the BMA framework. A uniform prior distribution for models takes each PS model as equally likely *a priori* and is considered a conservative choice that does not favour any particular specification. The Zellner prior for $\boldsymbol{\beta}_j$ (in symbols $\boldsymbol{\beta}_j \sim N(0, n(X_j'X_j)^{-1})$), takes the correlation among covariates (except for the constant, for which a diffuse prior is generally used) into account, differently from the often-used ridge prior. The intercept, moreover, is included in every model specification so, as a consequence, all other covariates are centred (Fernandez et al., 2001).

The model proposal distribution follows the one by Madigan et al. (1995), who propose to update the model specification at each step by simply adding or deleting one covariate. A more complex alternative which allows for more general movement patterns between models is also possible, but explorations in this direction made no significant difference. As for the differential function $f(\cdot)$ requested for the definition of new model parameters $\boldsymbol{\beta}_j$, we refer to Lucchetti and Pedini (2020), where the automatic RJMCMC sampler by Green (2003) and Lamnisos et al. (2009) is used. To sample models and parameters, we use 500000 MCMC iterations with a preliminary burn-in stage of 50000. Finally, the whole algorithm was performed using eight parallel Markov chains.

In Table 5 we report, for each variable, the posterior inclusion probability (PIP), the frequency with which a covariate is included in the PS specification, giving an indication

Table 4: Covariates main summary statistics (pre-treatment period)

	Full sample	treated	control
NEQU			
mean	1.700	1.973	1.635
std.dev.	0.570	0.548	0.555
median	1.500	2.000	1.500
min	1.000	1.000	1.000
max	4.400	3.800	4.400
CLY2			
category 1	0.172	0.075	0.195
category 2	0.184	0.155	0.191
category 3	0.188	0.206	0.184
category 4	0.215	0.267	0.202
category 5	0.241	0.297	0.228
AREA3			
North	0.416	0.471	0.403
Centre	0.196	0.206	0.194
South/Islands	0.388	0.323	0.403
ETA5			
≤ 30	0.025	0.051	0.019
$31 - 40$	0.095	0.198	0.070
$41 - 50$	0.195	0.339	0.161
$51 - 65$	0.323	0.337	0.320
> 65	0.362	0.075	0.430
STUDIO			
None	0.038	0.009	0.045
Elementary school	0.216	0.077	0.249
Middle school	0.353	0.455	0.328
High school	0.270	0.344	0.253
Bachelor degree	0.111	0.108	0.112
Post-graduate	0.012	0.007	0.013
QUALP10			
Blue-collar worker	0.149	0.359	0.099
Office worker	0.136	0.292	0.098
Junior manager	0.023	0.032	0.021
Manager	0.014	0.010	0.014
Member arts/professions	0.022	0.008	0.025
Sole proprietor	0.008	0.009	0.008
Freelance	0.041	0.027	0.044
Family business	0.008	0.003	0.009
Active shareholder	0.012	0.012	0.012
Not employed	0.589	0.248	0.670
CONDGEN			
Great difficulty	0.189	0.176	0.191
Difficulty	0.159	0.174	0.156
Some difficulty	0.300	0.340	0.291
Fairly easy	0.265	0.257	0.267
Easy	0.064	0.046	0.068
Very easy	0.023	0.007	0.026

of the variable's relevance. For the corresponding $\boldsymbol{\beta}$, we report the sample mean and standard deviation, in two varieties: unconditional (labelled "Mean" and "St. Err.") and conditional on variable inclusion (labelled "C. Mean" and "C. St. Err"); the latter are computed only on model specifications where the variable appears with a nonzero coefficient.

According to the PIP indicator, the number of household components and all the income classes seem to be extremely relevant, together with the South/Islands geographical dummy. As for the age categories, the PIP for all but the 31–40 interval are close to 1. For the education variable STUDIO, only the middle school and post-graduate categories report a PIP greater than 0.50. For the job type variable QUALP10, we have instead that most dummies exhibit a large inclusion frequency (with different intensity) except for office workers and sole proprietors. Among the economic deprivation dummies, the category with no perceived difficulties seem to show a significant effect.

In order to give the reader an idea of the impact of model uncertainty in the PS model specification, we report in Table 6 the estimation results for a probit model where all the covariates are used (Full) and for the 10 top-ranked probit models, from M1 to M10, that are the 10 models with the largest $\hat{\mathrm{P}}(M_j|\mathcal{D})$ from the RJMCMC sampler. We use ordinary frequentist estimation methods so as to adhere as closely as possible to the common practice in Propensity Score modelling. Although BMA should not be regarded as a model selection procedure, it is interesting to examine which models emerged with the highest posterior probability using, for example, information criteria, such as the BIC. Note that a complete scan of the model space for the lowest BIC would imply, in our case, analysing 2^{30} models, which is of course computationally infeasible.

Alternatively, a stepwise model selection procedure to identify the best models could be employed. However, this procedure may not be optimal, as some specifications may be skipped simply as a result of the sequential deletion/addition of variables. In this respect, the model space exploration provided by the RJMCMC grants a more widespread and computationally efficient analysis of the model set.

According to the posterior model probabilities, M1 appears to be slightly more probable than the others, with $\hat{\mathrm{P}}(M_j|\mathcal{D}) \approx 0.05$, which, interestingly, is associated to model M1 having the lowest BIC value among the top 10 selected.

5.2 ATET estimation

In this section, we compare the ATET estimates for our proposed BMA-based procedure with those yielded by the set of 10 top-ranked models selected according to the sampling frequency of the RJMCMC. Matching is performed using pairwise nearest neighbour with replacement, using 0.01 as caliper.

Table 7 reports the balancing properties resulting from the matching in the RJMCMC PS models. In order to assess the quality of the matching procedure, we perform paired and unpaired t-tests on the variables mean difference between treated and control units. As a summary measure of the test rejection rate, we use here the average p-value of the above tests, conditional on variable inclusion, over the 500000 RJMCMC samples;

Table 5: PS model: RJMCMC sampling statistics for the parameter vector $\boldsymbol{\beta}$

	Mean	St. Err.	PIP	C. Mean	C. St. Err.
Intercept	-1.114	0.029	1.000	-1.114	0.0291
NEQU	0.322	0.056	1.000	0.322	0.056
CLY2 (Ref: category 1)					
Category 2	0.398	0.107	0.986	0.404	0.096
Category 3	0.521	0.109	0.992	0.525	0.100
Category 4	0.553	0.117	0.992	0.557	0.106
Category 5	0.585	0.135	0.991	0.590	0.124
AREA3 (Ref: North)					
Center	0.000	0.0113	0.024	0.003	0.074
South/Islands	-0.130	0.091	0.739	-0.175	0.0573
ETA5 (Ref: $\leq$30)					
$31 - 40$	-0.014	0.072	0.075	-0.185	0.196
$41 - 50$	-0.299	0.111	0.964	-0.310	0.096
$51 - 65$	-0.544	0.104	1.000	-0.544	0.104
> 65	-0.999	0.126	1.000	-0.999	0.126
STUDIO (Ref: None)					
Elementary school	0.004	0.034	0.054	0.0796	0.123
Middle school	0.127	0.096	0.705	0.181	0.059
High school	-0.032	0.078	0.192	-0.165	0.099
Bachelor degree	-0.085	0.146	0.305	-0.279	0.126
Post-graduate	-0.379	0.425	0.512	-0.740	0.292
QUALP10 (Ref: Blue collar)					
Office worker	-0.002	0.022	0.033	-0.049	0.108
Jounior manager	-0.252	0.236	0.597	-0.423	0.146
Manager	-0.573	0.343	0.811	-0.707	0.224
Member arts/professions	-1.169	0.213	1.000	-1.169	0.213
Sole proprietor	-0.051	0.168	0.110	-0.465	0.253
Freelance	-0.889	0.129	1.000	-0.889	0.129
Family business	-1.294	0.330	0.999	-1.295	0.328
Active shareholder	-0.535	0.327	0.803	-0.667	0.214
Not employed	-0.764	0.067	1.000	-0.764	0.067
CONDGEN (Ref: Great difficulty)					
Difficulty	0.002	0.017	0.034	0.054	0.072
Some difficulty	0.007	0.030	0.085	0.088	0.061
Fairly easy	-0.038	0.076	0.238	-0.158	0.072
Easy	-0.090	0.150	0.313	-0.289	0.125
Very easy	-0.789	0.267	0.975	-0.809	0.238

Table 6: PS model: estimation results for the full and 10 top ranked Probit models (std. errors in parenthesis)

	Full	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
Intercept	-0.275	-0.963	-0.950	-0.981	-0.984	-0.883	-0.946	-0.895	-0.967	-0.751	-0.970
	(0.254)	(0.116)	(0.116)	(0.116)	(0.115)	(0.118)	(0.116)	(0.118)	(0.116)	(0.120)	(0.115)
NEQU	0.291	0.345	0.341	0.347	0.301	0.307	0.336	0.306	0.343	0.326	0.298
	(0.053)	(0.049)	(0.049)	(0.049)	(0.047)	(0.051)	(0.050)	(0.051)	(0.049)	(0.050)	(0.047)
CLY2 (Ref: Category 1)											
Category 2	0.433	0.398	0.401	0.396	0.407	0.412	0.401	0.411	0.398	0.405	0.410
	(0.099)	(0.096)	(0.096)	(0.096)	(0.095)	(0.096)	(0.096)	(0.096)	(0.096)	(0.095)	(0.095)
Category 3	0.606	0.507	0.512	0.499	0.531	0.549	0.513	0.544	0.505	0.540	0.535
	(0.102)	(0.094)	(0.094)	(0.094)	(0.093)	(0.095)	(0.094)	(0.095)	(0.094)	(0.094)	(0.093)
Category 4	0.670	0.518	0.524	0.509	0.579	0.594	0.531	0.591	0.515	0.569	0.583
	(0.108)	(0.094)	(0.094)	(0.094)	(0.091)	(0.097)	(0.094)	(0.096)	(0.094)	(0.095)	(0.091)
Category 5	0.790	0.532	0.547	0.503	0.613	0.675	0.568	0.659	0.520	0.617	0.627
	(0.122)	(0.099)	(0.099)	(0.099)	(0.095)	(0.108)	(0.100)	(0.107)	(0.099)	(0.102)	(0.095)
AREA3 (Ref: North)											
Center	-0.021	-	-	-	-	-	-	-	-	-	-
	(0.067)	-	-	-	-	-	-	-	-	-	-
South/Islands	-0.183	-0.171	-0.166	-0.172	-	-0.193	-0.176	-0.195	-0.167	-0.163	-
	(0.061)	(0.056)	(0.056)	(0.055)	-	(0.056)	(0.056)	(0.056)	(0.056)	(0.056)	-
ETA (Ref: $\leq$30)											
$31-40$	-0.232	-	-	-	-	-	-	-	-	-	-
	(0.144)	-	-	-	-	-	-	-	-	-	-
$41-50$ -0.498	-0.280	-0.285	-0.294	-0.278	-0.288	-0.282	-0.301	-0.300	-0.306	-0.284	
	(0.139)	(0.075)	(0.075)	(0.075)	(0.075)	(0.075)	(0.075)	(0.075)	(0.075)	(0.076)	(0.075)
$51-65$	-0.742	-0.513	-0.522	-0.521	-0.524	-0.517	-0.514	-0.525	-0.531	-0.562	-0.533
	(0.138)	(0.074)	(0.074)	(0.074)	(0.073)	(0.074)	(0.074)	(0.074)	(0.074)	(0.075)	(0.074)
> 65	-1.222	-0.954	-0.965	-0.956	-0.962	-0.959	-0.958	-0.961	-0.967	-1.059	-0.973
	(0.155)	(0.097)	(0.097)	(0.097)	(0.097)	(0.097)	(0.097)	(0.097)	(0.097)	(0.099)	(0.097)
STUDIO (Ref: None)											
Elementary school	-0.268	-	-	-	-	-	-	-	-	-	-
	(0.195)	-	-	-	-	-	-	-	-	-	-
Middle school	-0.194	0.174	0.168	0.191	0.187	0.171	0.173	0.186	0.185	-	0.181
	(0.195)	(0.052)	(0.052)	(0.052)	(0.052)	(0.052)	(0.052)	(0.052)	(0.052)	-	(0.052)
High school	-0.362	-	-	-	-	-	-	-	-	-0.180	-
	(0.199)	-	-	-	-	-	-	-	-	(0.059)	-
Bachelor's degree	-0.491	-	-	-	-	-	-	-	-	-0.320	-
	(0.211)	-	-	-	-	-	-	-	-	(0.090)	-
Post-graduate	-0.954	-	-0.579	-	-	-	-	-	-0.611	-0.807	-0.603
	(0.327)	-	(0.258)	-	-	-	-	-	(0.259)	(0.263)	(0.258)
QUALP10 (Ref: Blue collar)											
Office worker	-0.058	-	-	-	-	-	-	-	-	-	-
	(0.082)	-	-	-	-	-	-	-	-	-	-
Junior manager	-0.413	-0.435	-0.422	-	-0.439	-0.393	-0.418	-	-	-0.403	-0.424
	(0.152)	(0.141)	(0.141)	-	(0.141)	(0.141)	(0.141)	-	-	(0.141)	(0.141)
Manager	-0.668	-0.748	-0.678	-0.697	-0.770	-0.733	-0.743	-0.686	-0.627	-0.619	-0.697
	(0.228)	(0.210)	(0.215)	(0.210)	(0.210)	(0.212)	(0.211)	(0.212)	(0.215)	(0.217)	(0.215)
Member arts/prof.	-1.141	-1.218	-1.181	-1.174	-1.218	-1.181	-1.192	-1.139	-1.137	-1.127	-1.181
	(0.211)	(0.201)	(0.201)	(0.200)	(0.201)	(0.202)	(0.201)	(0.201)	(0.200)	(0.202)	(0.201)
Sole proprietor	-0.461	-	-	-	-	-	-	-	-	-	-
	(0.249)	-	-	-	-	-	-	-	-	-	-
Freelance	-0.937	-0.898	-0.902	-0.869	-0.900	-0.891	-0.897	-0.865	-0.874	-0.907	-0.903
	(0.134)	(0.127)	(0.127)	(0.127)	(0.127)	(0.127)	(0.127)	(0.127)	(0.127)	(0.128)	(0.127)
Family business	-1.346	-1.250	-1.261	-1.210	-1.257	-1.261	-1.257	-1.226	-1.223	-1.301	-1.269
	(0.321)	(0.313)	(0.314)	(0.313)	(0.314)	(0.317)	(0.315)	(0.317)	(0.313)	(0.315)	(0.315)
Active shareholder	-0.733	-0.672	-0.679	-0.633	-0.630	-0.680	-0.674	-0.645	-0.642	-0.679	-0.639
	(0.216)	(0.210)	(0.210)	(0.210)	(0.210)	(0.211)	(0.211)	(0.210)	(0.210)	(0.211)	(0.210)
Not employed	-0.822	-0.773	-0.775	-0.748	-0.781	-0.763	-0.766	-0.740	-0.751	-0.793	-0.782
	(0.073)	(0.063)	(0.063)	(0.063)	(0.063)	(0.063)	(0.063)	(0.063)	(0.063)	(0.064)	(0.063)
CONDGEN (Ref: Great difficulty)											
Difficulty	-0.013	-	-	-	-	-	-	-	-	-	-
	(0.087)	-	-	-	-	-	-	-	-	-	-
Some difficulty	-0.023	-	-	-	-	-	-	-	-	-	-
	(0.082)	-	-	-	-	-	-	-	-	-	-
Fairly easy	-0.188	-	-	-	-	-0.180	-	-0.192	-	-	-
	(0.094)	-	-	-	-	(0.065)	-	(0.065)	-	-	-
Easy	-0.328	-	-	-	-	-0.330	-0.232	-0.355	-	-	-
	(0.136)	-	-	-	-	(0.115)	(0.109)	(0.115)	-	-	-
Very easy	-0.859	-0.741	-0.753	-0.758	-0.739	-0.886	-0.776	-0.910	-0.770	-0.706	-0.751
	(0.240)	(0.221)	(0.222)	(0.222)	(0.221)	(0.225)	(0.222)	(0.226)	(0.223)	(0.224)	(0.222)
BIC	3633.0	3563.4	3566.2	3564.8	3564.5	3567.7	3567.1	3567.3	3567.0	3568.6	3566.8
$\hat{\mathrm{P}}(M \mid \mathcal{D})$	-	0.0487	0.0231	0.0188	0.0170	0.0165	0.0143	0.0138	0.0118	0.0105	0.01039

in other words, we average the p-values over all models that include that particular variable.

The PIP associated with each variable is reported in Table 5. As Table 7 shows, the null hypotheses of no mean differences in the covariates between the treated and control group, in both the paired and unpaired case, cannot be rejected.

The ATET estimator is computed in the same way as Neri et al. (2017): we take the difference between the 2014 and 2012 consumption ATET values, using the same matching group for both. This strategy produces an outcome equivalent to a difference-in-difference, where the comparison groups are adjusted for PS (sometimes known as propensity score diff-in-diff). Similar results could be obtained by taking as outcome variable the consumption difference between 2014 and 2012: even if the final ATET estimate will be the same, a different standard error will be produced. These results are not presented, but available upon requests.

Figure 1 shows the distributions of the ATET estimates across the model space for the five outcomes considered (see also Table 1 for a description). Each panel contains a kernel estimate of ATET density across the whole set of sampled models. All plots span a large range in ATET values. Null or negative impacts of the tax rebate can be found on the left tail of the densities for total, non durables, food and other durables consumption; the probability of such effects is even larger for car-related expenditure.

The final BMA estimates of the ATET are reported at the top of Table 8, where the averages across samples are computed as in Equation (5) and the associated standard error is evaluated as the square root of the variance in Equation (6). Notice that none of the ATET estimates are statistically significant. For comparison, the rest of the Table reports the ATET estimates based on the selected specification choices for the PS model, that are Full and M1 to M10. Test results for the related balancing conditions are available upon request.

Note that the results produced by models M1-M10 are quite heterogeneous. To be more specific, models M1, M2 and M6 produce a significant effect for total, non durables and food consumption, model M4 and M10 show significant ATET for food consumption while models M8 and M9 exhibit significant effects for non durables/food and total/non durables consumption, respectively.

The implication of Figure 1 and Table 8 is clear: blindly committing to model selection methods may lead to ignoring the fact that very similar specification, nearly equivalent in terms of fit and Information Criteria, may lead to dramatically different estimates of the ATET.

In order to further investigate the potential role of liquidity constraints, we estimate the effect of the tax rebate on a subset of households that are subject to financial strain and/or economic deprivation. Following Neri et al. (2017), two subsamples are selected. The first one is based on financial wealth, where a household is classified as liquidity constrained if its net financial wealth (difference between financial activities and liabilities) is smaller than half of its labour income (Broda and Parker, 2014). The second one embeds those families who perceive great difficulties to make ends meet, i.e. those belonging to the first category of the CONDGEN variable.

Table 7: Balancing conditions in BMA-RJMCMC: Paired and Unpaired t-tests p-values

	Unpaired p-value	Paired p-value
NEQU	0.62100	0.42821
CLY2 (Ref: Category 1)		
Category 2	0.53506	0.32832
Category 3	0.57794	0.38833
Category 4	0.53781	0.3228
Category 5	0.62168	0.41719
AREA3 (Ref: North)		
Center	0.37074	0.24227
South/Islands	0.51887	0.32847
ETA5 (Ref: $\leq$30)		
$31-40$	0.48597	0.32464
$41-50$	0.62956	0.43011
$51-65$	0.61098	0.36042
> 65	0.69072	0.35273
STUDIO (Ref: None)		
Elementary school	0.66068	0.48884
Middle school	0.40851	0.21517
High school	0.45373	0.26462
Bachelor's degree	0.51339	0.37247
Post-graduate	0.43066	0.41001
QUALP10 (Ref: Blue collar)		
Office worker	0.61926	0.42765
Junior manager	0.60105	0.48635
Manager	0.49527	0.42713
Member arts/professions	0.60515	0.48871
Sole proprietor	0.39222	0.38406
Freelance	0.59269	0.45853
Family business	0.58023	0.49427
Active shareholder	0.40669	0.37210
Not employed	0.75256	0.50859
CONDGEN (Ref: Great difficulty)		
Difficulty	0.22100	0.13715
Some difficulty	0.57952	0.41325
Fairly easy	0.62906	0.45421
Easy	0.55725	0.43848
Very easy	0.61658	0.50111

Table 8: ATET estimates: BMA, Full and 10 top-ranked models

	BIC	ATET				
		c_tot	c_food	c_notd	c_car	c_otherd
BMA		77.042	25.192	65.008	4.837	7.196
		(77.394)	(20.630)	(66.969)	(32.857)	(11.159)
Full	3633.0	59.398	23.519	51.373	13.297	-5.272
		(72.915)	(20.004)	(64.384)	(22.978)	(10.049)
M1	3563.4	102.826 *	33.083 **	90.360 *	2.329	10.137
		(61.661)	(15.803)	(54.292)	(20.734)	(8.594)
M2	3566.2	120.210 *	36.187 **	114.348 **	1.013	4.850
		(64.365)	(16.083)	(56.676)	(20.642)	(8.745)
M3	3564.8	70.292	25.430	64.482	5.898	-0.089
		(62.055)	(15.949)	(53.891)	(21.300)	(9.961)
M4	3564.5	80.933	29.035 *	87.173	-13.498	7.258
		(61.708)	(16.024)	(53.681)	(22.854)	(8.835)
M5	3567.7	91.222	27.845	60.156	29.010	2.056
		(63.273)	(17.215)	(54.031)	(23.628)	(8.955)
M6	3567.1	116.929 *	32.164 *	98.521 *	6.220	12.188
		(62.398)	(16.429)	(53.931)	(21.578)	(9.461)
M7	3567.3	69.808	17.924	43.890	14.364	11.554
		(64.815)	(17.383)	(57.261)	(21.579)	(9.147)
M8	3567.0	88.561	38.410 **	98.652 *	-11.562	1.472
		(63.614)	(16.510)	(55.961)	(21.027)	(8.502)
M9	3568.6	118.153 **	9.884	98.390 *	19.230	0.534
		(57.891)	(17.447)	(51.281)	(20.694)	(8.724)
M10	3566.8	44.476	34.488 **	72.549	-30.044	1.970
		(61.937)	(15.785)	(53.624)	(23.337)	(8.976)

*: p-value < 0.10, **: p-value < 0.05, ***: p-value < 0.01. Standard errors in parentheses.

Figure 1: ATET distributions across the model space

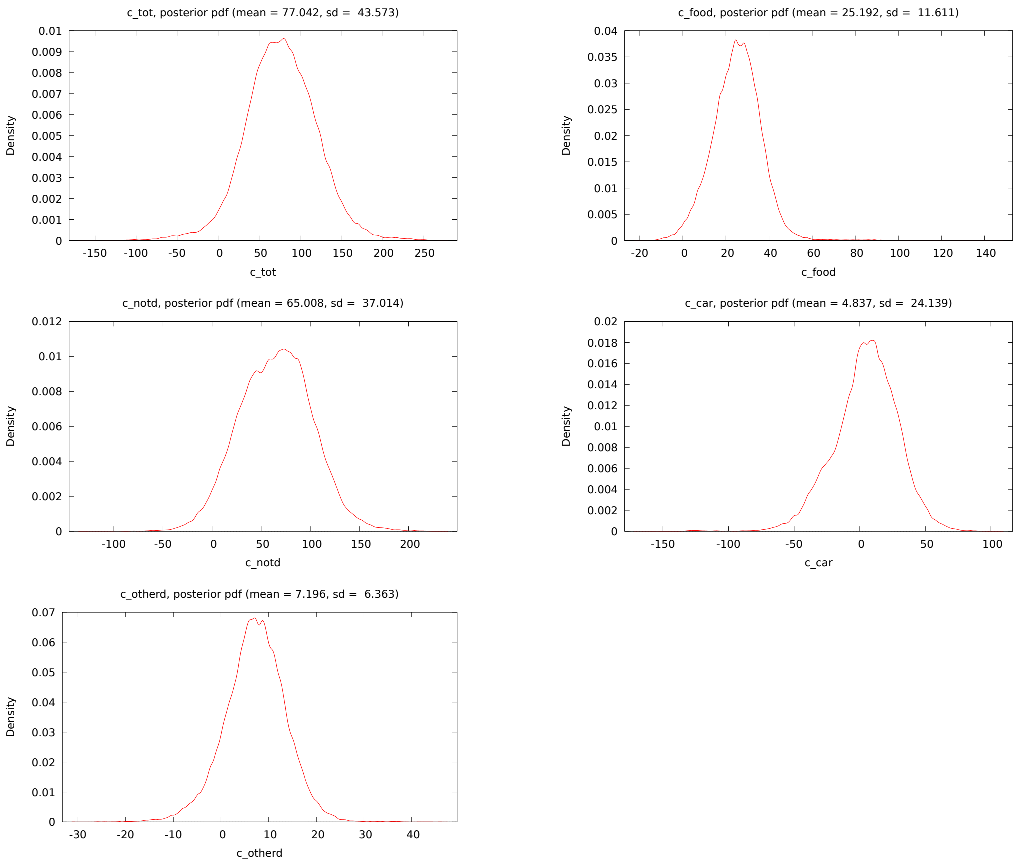

Table 9 reports the ATET estimates for the BMA-based PSM: the effect of the tax rebate on food consumption is statistically significant if only households facing economic deprivation are considered, which validates the liquidity constraint hypothesis. It is worth emphasising that a model selection procedure would have led us to conclude that the rebate was effective (see the ATET for food and non durables consumption using M1 in Table 8) for the whole sample of households, which would have been somewhat at odds with extant evidence in the related literature. Instead, model averaging coherently points toward a role for liquidity constraints, as the tax rebate has a significant effect on food expenditure only for those households struggling to make ends meet. This example effectively shows that model selection matters and how BMA makes it possible to handle model uncertainty, leading to a more complete portrait of the phenomenon and a more genuine evaluation of the ATET.

Table 9: BMA ATET estimates: liquidity constraints

	c_tot	c_food	c_notd	c_car	c_otherd
Financial wealth	9.422	23.397	3.217	6.193	0.012
	(76.538)	(23.129)	(62.675)	(34.451)	(12.627)
Economic deprivation	9.649	79.327**	10.137	-5.340	4.912
	(134.659)	(37.961)	(128.707)	(24.708)	(11.536)

*: p-value < 0.10, **: p-value < 0.05, ***: p-value < 0.01. Standard errors in parentheses. Financial wealth: $2,781$ households, 628 treated (23%) - RJMCMC with 100000 iterations after 10000 of burn-in. Economic deprivation: 841 households, 152 treated (18%). For the second subsample, the QUALP10 variable is expressed in three categories (employee, self-employed, not employed) for identification purposes as well as the post-graduate educational level class is embodied in the Bachelor's degree category. RJMCMC with 500000 iterations after 50000 of burn-in.

6 Discussion and conclusions

Model uncertainty is an impelling problem in any model building procedure, but little attention to the topic is devoted in PSM literature. In this article, we propose a BMA solution based on RJMCMC for the PSM estimation of treatment effects. Using the SHIW dataset, we estimate the consumption effects of the Italian 2014 tax rebate.

Results show that the impact of the tax rebate on consumption is quite different across the expenditure categories considered. In fact, the additional consumption is mainly driven by an increase in the expenditure for non durable goods. This result is in line with that of Neri et al. (2017), who report similar findings for the self-reported spending of the tax rebate in the SHIW survey. However, such increase in expenditure is statistically insignificant for all of the consumption categories considered here.

Perhaps more interestingly, the problem of model uncertainty is evident in our application. By using a model selection procedure, we show that slightly different choices for the PS specification can lead to markedly different estimates of the ATET. Consider, for instance, the estimated ATET resulting from the PS model with the smallest BIC (see Section 5.1). Judging from these, the effects on total, non durables and food consumption turn out to be statistically significant. Such results may enable the conclusion that tax credit is indeed an effective policy instrument in contrast to the consumption smoothing theory. However, after taking model uncertainty into account, this positive effect is called into question, as the BMA-based PSM estimates point toward statistically insignificant treatment effects for every consumption category. If any, model averaging suggests that the tax rebate had an effect on food expenditure for liquidity constrained households, in line with previous findings.

References

Agarwal, S., C. Liu, and N. S. Souleles (2007): "The reaction of consumer spending and debt to tax rebates. Evidence from consumer credit data," *Journal of political Economy*, 115, 986–1019.

An, W. (2010): "Bayesian Propensity Score Estimators: Incorporating Uncertainties in Propensity Scores into Causal Inference," *Sociological Methodology*, 40, 151–189.

Belloni, A., V. Chernozhukov, I. Fernández-Val, and C. Hansen (2017): "Program evaluation and causal inference with high-dimensional data," *Econometrica*, 85, 233–298.

Belloni, A., V. Chernozhukov, and C. Hansen (2014): "Inference on treatment effects after selection among high-dimensional controls," *The Review of Economic Studies*, 81, 608–650.

Broda, C. and J. A. Parker (2014): "The economic stimulus payments of 2008 and the aggregate demand for consumption," *Journal of Monetary Economics*, 68, S20–S36.

Brookhart, M. A., S. Schneeweiss, K. J. Rothman, R. J. Glynn, J. Avorn, and T. Stürmer (2006): "Variable selection for propensity score models," *American journal of epidemiology*, 163, 1149–1156.

Bryson, A., R. Dorsett, and S. Purdon (2002): "The use of propensity score matching in the evaluation of active labour market policies," *LSE Department for Work and Pensions Working Paper*.

Caliendo, M. and S. Kopeinig (2008): "Some practical guidance for the implementation of propensity score matching," *Journal of economic surveys*, 22, 31–72.

Cefalu, M., F. Dominici, N. Arvold, and G. Parmigiani (2017): "Model averaged double robust estimation," *Biometrics*, 73, 410–421.

Chatfield, C. (1995): "Model Uncertainty, Data Mining and Statistical Inference," *Journal of the Royal Statistical Society*, 158, 419–466.

Chen, M.-H., L. Huang, J. G. Ibrahim, and S. Kim (2008): "Bayesian Variable Selection and Computation for Generalized Linear Models with Conjugate Priors," *Bayesian Analysis*, 3, 585–614.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018): "Double/debiased machine learning for treatment and structural parameters," *The Econometrics Journal*, 21, C1–C68.

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.

Fernandez, C., E. Ley, and M. F. Steel (2001): "Benchmark priors for Bayesian model averaging," *Journal of Econometrics*, 100, 381–427.

George, E. I. and R. E. McCulloch (1993): "Variable selection via Gibbs sampling," *Journal of the American Statistical Association*, 88, 881–889.

Green, P. J. (1995): "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, 82, 711–732.

——— (2003): "Trans-dimensional Markov Chain Monte Carlo," *Oxford Statistical Science Series*, 179–198.

Hastings, W. K. (1970): "Monte Carlo sampling methods using Markov Chains and their applications," *Biometrika*, 57, 97–109.

Heckman, J. J., H. Ichimura, and P. E. Todd (1997): "Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme," *The review of economic studies*, 64, 605–654.

Heim, B. T. (2007): "The effect of tax rebates on consumption expenditures: evidence from state tax rebates," *National Tax Journal*, 685–710.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999): "Bayesian Model Averaging: a tutorial," *Statistical science*, 382–401.

Holmes, C. C. and L. Held (2006): "Bayesian auxiliary variable models for binary and multinomial regression," *Bayesian analysis*, 1, 145–168.

Jappelli, T. and L. Pistaferri (2010): "The consumption response to income changes," *Annual Reviews*, 2.

Johnson, D. S., J. A. Parker, and N. S. Souleles (2006): "Household expenditure and the income tax rebates of 2001," *American Economic Review*, 96, 1589–1610.

Kaplan, D. and J. Chen (2012): "A two-step Bayesian approach for propensity score analysis: Simulations and case study," *Psychometrika*, 77, 581–609.

——— (2014): "Bayesian model averaging for propensity score analysis," *Multivariate behavioral research*, 49, 505–517.

Kitagawa, T. and C. Muris (2016): "Model averaging in semiparametric estimation of treatment effects," *Journal of Econometrics*, 193, 271–289.

Lamnisos, D., J. E. Griffin, and M. F. Steel (2009): "Transdimensional sampling algorithms for Bayesian variable selection in classification problems with many more variables than observations," *Journal of Computational and Graphical Statistics*, 18, 592–612.

Lucchetti, R. and L. Pedini (2020): "ParMA: parallelised Bayesian Model Averaging for Generalised Linear Models," Tech. Rep. 28/WP/2020, Ca Foscari University of Venice, Department of Economics Working Paper.

Madigan, D., J. York, and D. Allard (1995): "Bayesian graphical models for discrete data," *International Statistical Review/Revue Internationale de Statistique*, 215–232.

Millimet, D. L. and R. Tchernis (2009): "On the specification of propensity scores, with applications to the analysis of trade policies," *Journal of Business & Economic Statistics*, 27, 397–415.

Neri, A., C. Rondinelli, and F. Scoccianti (2017): "Household spending out of a tax rebate: Italian "€80 tax bonus"," Tech. Rep. 379, Bank of Italy (Occasional Papers).

Parker, J. A. (1999): "The reaction of household consumption to predictable changes in social security taxes," *American Economic Review*, 89, 959–973.

Parker, J. A., N. S. Souleles, D. S. Johnson, and R. McClelland (2013): "Consumer spending and the economic stimulus payments of 2008," *American Economic Review*, 103, 2530–53.

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

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

Rubin, D. B. and N. Thomas (1996): "Matching using estimated propensity scores: relating theory to practice," *Biometrics*, 249–264.

Shapiro, M. D. and J. Slemrod (2003a): "Consumer response to tax rebates," *American Economic Review*, 93, 381–396.

——— (2003b): "Did the 2001 tax rebate stimulate spending? Evidence from taxpayer surveys," *Tax policy and the economy*, 17, 83–109.

——— (2009): "Did the 2008 tax rebates stimulate spending?" *American Economic Review*, 99, 374–79.

Souleles, N. S. (1999): "The response of household consumption to income tax refunds," *American Economic Review*, 89, 947–958.

——— (2002): "Consumer response to the Reagan tax cuts," *Journal of Public Economics*, 85, 99–120.

Steel, M. F. (2020): "Model averaging and its use in economics," *Journal of Economic Literature*, 58, 644–719.

Wilcox, D. W. (1989): "Social security benefits, consumption expenditure, and the life cycle hypothesis," *Journal of Political Economy*, 97, 288–304.

Xie, Y., Y. Zhu, C. A. Cotton, and P. Wu (2019): "A model averaging approach for estimating propensity scores by optimizing balance," *Statistical methods in medical research*, 28, 84–101.

Zellner, A. and A. Siow (1980): "Posterior odds ratios for selected regression hypotheses," *Trabajos de estadística y de investigación operativa*, 31, 585–603.

Zigler, C. M. and F. Dominici (2014): "Uncertainty in propensity score estimation: Bayesian methods for variable selection and model-averaged causal effects," *Journal of the American Statistical Association*, 109, 95–107.