Synthetic Control Method versus Standard Statistical Techniques: a Comparison for Labor Market Reforms

Research · November 2015

DOI: 10.13140/RG.2.1.1301.4486

1 author:

Rafael Valero
Keele University

Synthetic Control Method versus Standard Statistical Techniques: a Comparison for Labor Market Reforms

Rafael Valero *

November 7, 2015

Abstract

The effect of structural reforms is key when evaluating policies. The difficulty is that once the reform has been implemented it is not possible to observe the counterfactual, i.e. what would have happened if the reform had not occurred. The need of an evaluation is very important, yet in many cases not possible, such as for different scenarios in labor market reforms; something at the forefront of the minds of policy makers in their aim to reduce unemployment. The current literature is dominated by the Synthetic Control Method (SCM), but there are some shortcomings of this method. Firstly, there are cases where it cannot be performed; and secondly, it was not created considering the trade-off between bias and variance, hence the accuracy of results could be improved. I have modified the current method to alleviate the SCM weaknesses. The main contribution is to propose standard statistical alternatives to the SCM; with the main focus on LASSO with cross-validation. The alternatives are widely studied and implemented in other sciences and inbuilt in most software packages. A comparison between the pair of methodologies is provided; both theoretically and by using different tests in real data sets. The empirical contribution is to apply the previous methods to two labor market reforms: a) for a reduction of the workweek load, I study Portugal (1996); and b) for a reduction of firing cost, I study Spain (2010). The results show that there is a reduction of unemployment in Portugal. For the Spanish case the result showed an increase in unemployment rate.

JEL classification : C18, E24, J08, J21, J22

Key Words : policy evaluation, counterfactual series methods, labor economics

*University of Alincante, Department of Economics: rafael.valero@ua.es or rafael.valero.fernandez@gmail.com. The author is extremely grateful to Gabriel Perez-Quiros for discussion, comments, suggestions and time. I thank Pedro Albarrán, María Ángeles Carnero, Lola Collado, Lilia Maliar, Serguei Maliar, Asier Mariscal, Juan Mora, Francesco Serti and seminar participants at University of Alicante for their helpful comments and suggestions. My sincere thanks to Rebecca L. Gidman who has provided immense help with preparation. I also acknowledge financial support from MECD under the FPU program. The author will provide the software. Errors are mine.

1 Introduction

The object of study of an economist cannot usually be implemented as a controlled experiment to gather data and test theories, in contrast with other disciplines. However by using statistics, in some cases, conclusions might be found.

Imagine that a change happens, e.g. a change in law or a hurricane. Some important questions are: What would have happened if there had not been a change? Would there be any difference in a relevant variable? and if so, In what direction? How much? and so on and so forth: in other words, the evaluation of a policy or another event. To answer these questions by using Counterfactual Series, i.e: artificial series that have not happened but would if there had not been a change, the cutting edge methodology is the Synthetic Control Method (SCM, introduced by Abadie and Gardeazabal 2004 and developed by Abadie, Diamond and Hainmueller 2010, hereafter ADH 2010).

The SCM has at least three limitations: firstly, it is impossible, by construction to use it in some cases. Secondly, it was designed without considering the trade-off between bias and variance; and thirdly, there are difficulties guaranteeing that there was a change and the counterfactual series is reliable. All these limitations are treated here, from different angles: the first one by adding a constant and the second by using Cross-Validation[1] but with limited success, this is a contribution of this paper. By changing the methodology[2], all the shortcoming are simultaneously addressed with better outcome.

The main contribution of this paper is to use, in this framework, other standard statistical techniques which perform remarkably well, without the previously mentioned limitations. Broadly used for other scientific fields[3], well studied and implemented in most of the statistical software packages[4], I will focus in LASSO (Tibshirani 1996) implemented with Cross-Validation (CV). I do this for two reasons: a) it is possible to prove that SCM is a particular case of LASSO, and b) because some variables may be ruled out, making the interpretation simpler and easier, given that to trust the counterfactual series, the behavior of all variables must be taken into consideration. Alternative techniques are pointed out, Counterfactual Series Methods are those methods which allow the creation of counterfactual series.

For further assessment I have implemented different tests. Therefore theoretical evidence (in general, in terms of minimization of squared errors, predictive ability and the cases where it is applicable) and (in particular) empirical tests, are provided to compare the methodologies: another contribution of this paper.

The empirical contribution of this paper includes the employment of both techniques,

[1] By means of "Jackknife model averaging" by Hansen and Racine (2012). The software will be available.

[2] The third limitation comes, in part, as a consequence of the second limitation, check that effectively a change has happened and that the counterfactual series is reliable (by ensuring that the other units are not changing too much at the same time), using a placebo test.

[3] Probably Finances is the closest.

[4] I follow Hastie et al. 2009, which is probably the most important general reference for the topic.

i.e. the SCM and LASSO with Cross-Validation in real cases. The real empirical cases are two different kinds of Labor Market reforms: a) a reduction of maximum work week hours, there example of Portugal in 1996; and b) the reduction of firing cost, the example is Spain in 2010. In previous works there is no strong evidence that those types of labor reforms were effective or have a macroeconomic impact; further literature about this kind of labor market reform is provided in Section 3. The results are that Portugal enjoyed a reduction of unemployment. For the Spanish case the law showed an apparent increase in unemployment.

The structure of this paper is as follows: in section 2, the methodology is described. In Section 3 there is a literature review of the treated points and countries. Section 4 depicts the results and Section 5 the conclusion. There are further appendices which complement and detail particular aspects, such as data description, Section 5; placebo tests, Section 7; tests to verify the quality of the models, Section 9; an explanation of the trade off between bias and variance and model selection, Section 10; graphical examples of the modifications proposed for SCM, Section 8; and the coefficients for the particular cases, Section 6.

2 Methodology

In this section the terminology is clarified, some definitions are introduced and the methodology is justified and explained in comparative terms. The main motivations are that SCM cannot perform well in all cases, for different concerns which are detailed later, probably the most important being the bias-variance trade-off and the selection of the model complexity. The SCM is the most extended methodology in the field, but it was not designed for the previous motivation points, in fact, the SCM could be enhanced by using some of the techniques explained here, for example Cross-Validation. Terminology is clarified, from the SCM and statistical learning.

The main methodological conclusion is that LASSO combined with Cross-Validation appears as a solid option, given that: 1) it could be seen as a generalization of the SCM, 2) it can perform naturally in more cases and 3) probably more accurately. In order to show that, theoretical and practical evidence is presented. For practical evidence different tests are proposed to be used in real datasets.

By using the pretreatment data, a model is estimated; after that, the estimates are used to create a counterfactual series and this series is compared with the real one and then conclusions can be reached.

It is not enough to use the methodology proposed before, but check that no strong changes happen in the other units when the counterfactual series is created, otherwise the counterfactual series may be affected by other effects different to the one object of the study. In order to check for that a way is, as explained in ADH 2010, using the methodology but in the rest of the units, with the same breakpoint, and observe that the unit which suffered the largest effect is the treated one, it is a placebo tests.

2.1 Terminology, notations and examples

Units are (as in Abadie et al. 2010 and experimental literature) the subject to study, for example countries, regions, firms or cities. The units can be divided into a treated unit; the one which suffers the change of interest, and the control units; which are unaffected and similar to the treated unit. Control units are the selected ones, and after that, they are used to create a counterfactual/synthetic series which the one to compare with (the authentic control).

The breakpoint or treatment, i.e. the initial date of a change, is, for example, the introduction of a new law, the inclusion of a country to the European Union or an hurricane.

In those terms and including the definitions stated in the Introduction, i.e. counterfactual series and CSM, it is possible to keep in mind the next example, which will be studied with detail later. The aim is to create a counterfactual series for unemployment in Portugal. Portugal implemented a law in December 1996, to reduce the maximum standard workweek to 40, from, in some cases 44; the law was implemented within two years and two steps. Unemployment data is gathered for all OECD countries and applied to one or more CSM. The units are the countries of the OECD area, and the treated unit is Portugal. To create the counterfactual series the rest of the units are used. Models are estimated before the treatment, and the estimates are used to create the counterfactual series after the treatment and by comparing the real series with the counterfactual series it is possible to draw some conclusions.

In SCM certain measures called predictors are also used, they covariate with the reference variable, which are supposed to help the selection of units, see Abadie et al. 2010. For example given that unemployment is the topic here, some predictors could be the average education of the work force, the average of the proportion of women, the unemployment age composition or unemployment duration. Following Abadie et al. 2010, $J+1$ is the number of units, T the number of periods, T_0 the number of pre-treatment periods, R the number of covariates or predictors.

In this way, it is useful to define the next vectors: $Y_1 \equiv$ vector of values of the interesting series of the treated unit, $1 \times T$, $Y_0 \equiv$ the same for the rest of the units, $J \times T$, $Z_1 \equiv$ vector of pre-treatment values of the interesting series of the treated unit, $J \times T_0$,

$Z_0 \equiv$ the same for the rest of the units, $1 \times T_0$, $X_1 \equiv$ vector of pre-treatment values of R predictors for the treated unit, $J \times R$, $X_0 \equiv$ the same for the rest of the units, $1 \times R$.

From statistical learning: the data set can be split in training, validation and test sets. The training set is the set where the model is fitted, the validation set is where the prediction error is obtained and the models selected and the test set is where finally the selected model is assessed[5]. In contrast with the previous paragraph predictors or inputs are the independent variables.

2.2 SCM and Pseudo SCM

In this section I explain how to create counterfactual series by using SCM. SCM, developed in Abadie and Gardeazabal 2003 and Abadie, Diamond and Hainmueller 2010, what could be written as follows:

$$V = \arg\min_{V \epsilon \vee} (Z_1 - Z_0 W^*(V))'(Z_1 - Z_0 W^*(V)) \tag{1}$$

s.t.:

$$W = \arg\min_{W \epsilon \omega} (X_1 - X_0 W)' V (X_1 - X_0 W) \tag{2}$$

s.t:

$$\sum_{i=2}^{J+1} w_i = 1 \tag{3}$$

$$0 \leq w_i \leq 1, i = 2, ..., J+1 \tag{4}$$

where $\vee$ is the set of all nonnegative diagonal matrices, $W^*(V)$ is the solution from 2 constraint to 3 and 4, and ω is a set of vectors such as $[w_2, w_3, ..., w_{J+1}]$. See Abadie and Gardeazabal, (2003), Appendix B. In the same appendix, they also proposed an alternative way, it is referred to as pseudo SCM for simplicity:

$$\min_{W} (Z_1 - Z_0 W)'(Z_1 - Z_0 W) \tag{5}$$

such that constraints 3 and 4 are met. The authors pointed out that pseudo SCM (PSCM) could be less appropriate since it does not consider more information about the units, i.e. X_0 and X_1.

Lemma 1 *MSE of pre-treatment periods is equal or lower for PSCM than SCM.*

Proof. Because there is a constraint less for pseudo SCM. ■

The authors introduce the constraints 3 and 4 SCM and PSCM with the idea of avoiding extrapolation. "Synthetic control is a weighted average of the available control units" Abadie et al. 2010, this interpretation is very clear and intuitive. The idea behind is to select similar units, therefore they should behave in a similar way.

After coefficients W are selected post-treatment data is used to create the counterfactual series. The real and the counterfactual series can be compared to draw conclusions.

[5] Use here to comepare among PSCM and LASSO-CV.

SCM has been used in a variety of topics, such as terrorism (Abadie and Gardeazabal 2004), different sorts of catastrophes (Cavallo et al. 2013) or CO2 emissions (Almer and Winkler, 2012). It is currently a very used quantitative approach.

To create the counterfactual series the coefficient are used Y_0W, this is the counterfactual serie and the real is Y_1.

2.2.1 Shortcomings

By construction, there are at least three:

1. It is not possible to construct a counterfactual series for units which are always at the top of at the bottom to the study series, i.e. Spain in terms of unemployment is a top country among the OECD countries, PCSM is used, see Figure 5.
2. The trade off between bias and variance is not considered[6].
3. Placebo tests may not pass because of the inability of SCM, i.e. it may be a real change but the methodology is not able of notice it.

The first point may be alleviated by introducing a constant. The second shortcoming is more difficult to overcome, a solution is also implemented. Both cases are explained in Section 8. The third shortcoming is alleviated by using other methods.

2.3 Standard Statistical Techniques

Given the previous shortcomings it seems worthy to implement techniques which help us to overcome them, with this idea in mind, standard statistical techniques are introduced here, mostly following Statistical Learning. From all the possibilities[7] LASSO is chosen in order to compare directly with SCM or PSCM, given that both can be seen as a restricted LASSO case. For further discussion see Hastie et al. 2009, particularly Ch. 3 and 7. It is possible to split this section in two, for the model, LASSO is chosen, and for the model selection procedure, Cross-Validation is chosen.

2.3.1 Linear Models

By restricting the model to only linear there are a few of them, such as Subset Selection, Ridge Regression, Principal Component Regression, Partial least Squares, LASSO or LARS, among others. For the purpose of this paper LASSO, LARS and Ridge Regression may work under some conditions. See Hastie et al. 2009, particularly Ch. 3.

[6] Detail explanation is provided in a general case in an appendix.

[7] There is a large variety of models. Only for linear cases: Subset Selection, Ridge Regression, Principal Component Regression or Partial Least Square, at least.

LASSO Tibshirani 1996 created a procedure to combine the reduction of variables of Subset Selection (SS) and the smoothness of Ridge Regression. LASSO, Least Absolute Shrinkage and Selection Operator was the result. LASSO is an appropriate fit, it is able to select units in a softer way than SS, is more extended and is broadly implemented in software packages. For the purposes of this paper only pre-treatment variables are introduced, i.e. Z_1 and Z_0.

With the previous notation, it can be written as:

$$\min_W (Z_1 - c - Z_0W)'(Z_1 - c - Z_0W) \tag{6}$$

s.t.:

$$\sum\nolimits_2^{J+1} |w_i| \leq t \tag{7}$$

where $t \succeq 0$ is a tuning parameter which controls the amount of coefficient shrinkage and c is an intercept and does not account for the constraint 7. To provide some insight, consider the OLS estimation coefficients $\hat{w}^{OLS}$, suppose $t_0 = \sum_2^{J+1} \left|\hat{w}_i^{OLS}\right|$, then a $t = t_0\alpha$, suppose there is an reduction $(0 \leq 1 - \alpha \leq 1)$ of $\alpha\%$ on average of the OLS estimated coefficients.

Lemma 2 *MSE of pre-treatment periods is equal or lower to LASSO with* $t \succeq 1$ *than for PSCM and SCM.*

Proof. Constraint 7 covers more cases than constraints 2, 3 and 4. ■

With respect to PSCM and SCM, LASSO allows for negative coefficients that may not have interpretation, however they provide counterfactual series with some good properties.

From all previous techniques and for all previous reasons, LASSO is the one which better matches the purposes of this paper. 1) It can be shown as a more broad model than PSCM or SCM, 2) it has been extensively used for other purposes, 3) some units may be ruled out, making both interpretations easier, and the worry of changes in theses units reduced.

Shortcomings Possible LASSO shortcomings are mainly two and possible solutions are enclosed also; a) the selection of the appropriate complexity parameter t, see Model Selection Section 2.3.2, b) in different version of the software the constant term is not used to compute t, meaning there is no penalty of the constant term, therefore the selection of the constant term maybe larger, however by normalizing the data set is the constant term may be removed.

2.3.2 Model selection

Model selection is based on the selection of a model accordingly, not only with its ability to minimize square errors but also its ability to predict. There are different ways to do it (see Chapter 7 from Hastie et al. 2009 for a discussion). There are analytical approaches, such as Akaike information criterion (AIC); Bayesian information criterion (BIC); minimum description length (MDL) and structural risk minimization (SRM). And numerical approaches, such as Holdout method; Cross-Validation (CV); or Bootstrap. For a comparison among the last three see Kohavi 1995.

I focus on Cross-Validation, given its simplicity, broad uses and its implementations in software packages, acknowledging that there are others and there are also different CV procedures.

Cross Validation The easiest CV kind to understand is probably k-fold CV. The basic idea could be summarized briefly as follows, but it may change according to the software used:

1. Split the initial data set in a number of sets K.
2. For $i = 1, ..., K$
   - Use the remaining $K - 1$ to estimate the model or models (training, these subsets form the training set).
   - Evaluate in the set i (called validation or testing set) the previous estimation or estimations.
   - Combine the estimate or estimates, for example by taking averages.
3. Select the best model.

Alternatively Monte Carlo could be implemented by repeating 1 (with random splits) and 2 many times. There are more complicated versions, but this is fairly easy to explain; in software packages there may be a different kind, but the idea is roughly the same. The method is sensitive to the number K, given that it determines the size of the training sets. Usual K values are 5 or 10. For further discussion on CV see Arlot and Celisse 2010. Therefore when in LASSO t is chosen is for a reason, it is supposed to be the best option. A similar algorithm is implemented for SCM and PSCM to avoid the trade-off between bias and variance, as much as possible, but with limited success. The "Jackknife model averaging" by Hansen and Racine (2012) is implemented, but in this case to select the best coefficients not a (complexity) parameter, such a t for LASSO, therefore the impact may be less noticeable.

2.4 Implementation and Robustness

In order to trust the final counterfactual series, in line with the literature of SCM, a placebo test is used; where keeping the breakpoint constant, each unit is analyzed, to be sure that the treated unit is effectively suffering a change, and that the non treated units are not having strong idiosyncratic shocks, which may affect the results.

The same idea may be done in different ways, the standard ones are provided and other with the same spirit but different execution, details is Section 7.

2.5 PSCM vs Standard Statistical tools a summary

Previously it has been shown some shortcoming has been shown for SCM and PSCM, I have try to overcome these by modifying the SCM and PSCM algorithm; apparently the introduction of a constant can be more helpful than including Cross-Validation. All in all, these modifications have apparently limited impact, see Section 8 for a particular example. When standard statistical techniques are used, they can overcome the shortcomings. From all different combinations I choose LASSO and Cross-Validation[8].

The pool of unit should be carefully chosen and placebo tests should be run in order to have guarantees of the existence of a breakpoint and the non treated units do not have, in the meanwhile, important idiosyncratic shocks; for these purposes placebo tests are implemented.

LASSO can be seen as a generalization of SCM, when t is equal or larger than 1, since t is chosen with a purpose, in general LASSO with Cross-Validation may perform better. The researchers can freely choose how to do it according to their particular cases and if they accept the restrictions and possible critiques which come with each methodology. There are a number of tests to choose among SCM, PSCM, LASSO-CV or others, which may help for the particular cases.

To create these tests a number of different subsets from the pretreatment dataset are reserved only for assessment and the average and standard difference is reported. This is the spirit of testing in Statistical Learning. Two very simple tests are implemented here 1) The Random Points Test and 2) The Windows test, which are applied and explained in Section 9.

3 Labor Economics Applications

Changes in labor laws are considered: Firstly, three cases of reduction of maximum standard workweek: Portugal (1996). Secondly, a case of a decreases in firing costs: Spain (2010).

3.1 Short Literature Review

In labor economics there are many studies on the effect of different laws, laws because, usually, all of them are different, in content or implementation or enforcement, although

[8] There are alternatives for both, as explained before.

the objective could be the same. For instance, the change in the law for South Korea was implemented in six steps from 2004 until 2011, or the prototypal example of France, where after the implementation in 2000 changes have been introduced. Therefore they are not ideal conditions for researching. There are very weak macroeconomic results in general. The literature is vast and complex, given not only the different laws, but the different available data and the very different cases that should be faced. Conclusion at a microeconomic level could affect married couples in a different way than single individuals (see Askenazy 2003, Rudolf and Seo-Young 2011). If employees are highly qualified, they would go directly to extra hours (see Askenazy 2003) and therefore they earn more money; or people who are already employees could look for secondary part time jobs (see Askenazy 2003) and so on and so forth. If unions are very strong they claim full salaries and this may lead to maintaining the same salary, undermining possible job creation (see Hunt 1998, for Germany, Raposo and van Ours 2010, for Portugal, Sánchez 2010, for Chile). Disentangling all previous effects at a macroeconomic level could be dubious[9]. In general, evidence of any sort of change is not found in the long term (see Jacobson and Ohlsson 2000 for Sweden, and Kapteyn et al. 2003 for OECD countries). I refer to Askenazy not only as a literature review of the French case for the 2000 and posterior years, but a general reference for reduction of maximal standard workweek hours, given the importance and the amount of research for the French case.

For the second empirical case in this study, Spain in 2010, the law was oriented to reduced cost of firing. Within the literature there is at least one reference which did not find evidence for any effect (see Hunt 2000, in this case for Germany in 1985). The year and country is written in order to illustrate the difficulties in comparison, for instance comparing 2010 Spain's reality with 1985 Germany's, which is very a common difficulty in Comparative Politics.

4 Results

4.1 Implementation

Data comes from OCDE. All units are in percentages; for further description see Section 5.

As pointed out in Abadie et al. 2014, all units should have similar characteristics and units with similar treatment or with strong idiosyncratic shocks should be excluded, although it depends on when they happened. If it happened before the period of our interest it is probably not going to affect the results, given that the methods account for it, it is always possible to perform a secondary analysis excluding the suspicious series. Greece is removed given their particular situation since 2008, also a placebo test is realized to

[9] Maybe because of the difficulties of comparing countries with previous tecqniques, however studies based on SCM with countries as units such as Abadie et al. 2014, Billmeier and Nannicini 2013 or Cavallo et al. 2013 have been employed for researching in other areas.

removed those units with large idiosyncratic shocks. The breakpoints have been analyzed by using placebo tests in Section 7, in order to guarantee or not the breakpoints' existence and reliability of the counterfactual series.

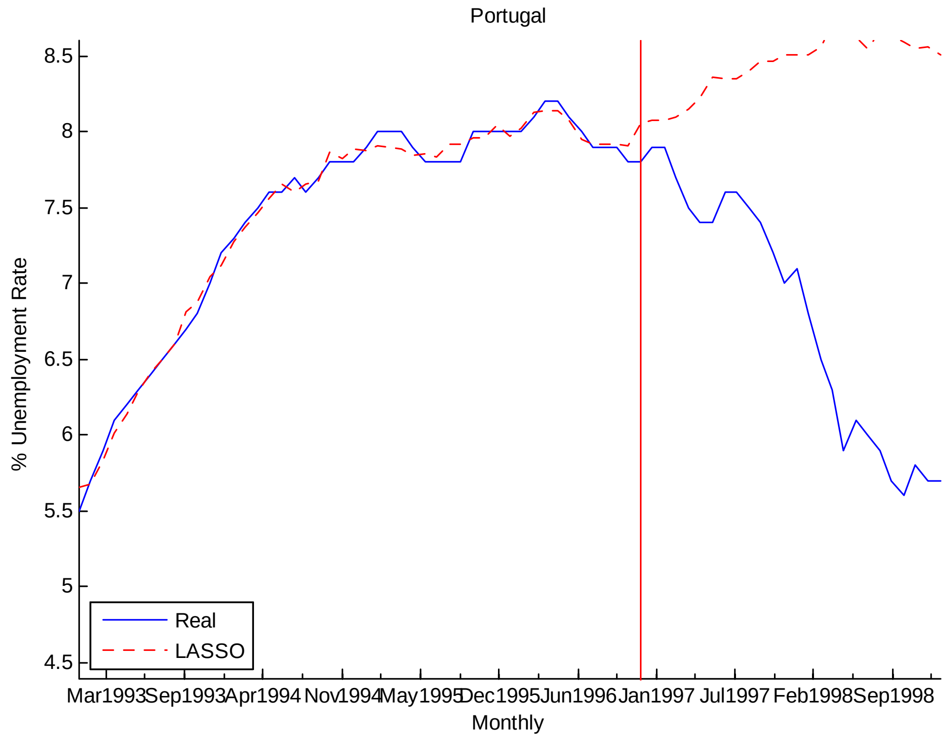
A brief description of the law and previous works on the same changes are provided.

4.2 Portugal

On the first of December 1996, by law the maximum standard workweek in Portugal was reduced from 44 hours to 40 in two steps; firstly to 42 (those between 42 and 40, to 40), and then by 1997 all of them had changed to 40. As Raposo and Van Our (2010) point out, the main purpose of the law was not to create new jobs or reduce unemployment, but to converge to the European standards. By examining individual data, Raposo and Van Our (2010) found that for the workers involved, this change reduced the job separation rate and increased hourly wages, keeping monthly earnings approximately constant. The working hour reduction also affected those working less than 40 hours per week; they were more likely to lose their job.

As seen in Figure 1, there is a reduction in unemployment with respect to LASSO-CV. PSCM does not pass placebo tests therefore its use for the case is not appropriate, although in Section 6 the coefficients and the graphs are displayed. According to Figure 1 LASSO-CV is a good fit. Apparently the law started to have macroeconomic effect soon and some employment was created, roughly around a 2% of reduction on unemployment. It is reported until 1999 because Portugal joined the euro at this time; therefore it is possible that the "euro effect" started before 1999. A possible shortcoming here is there is not to much pretreatment data.

Figure 1



As described before, tests are computed in order to compare both methodologies, for the "Random Points Test" a 10% of the pretreatment sample is always saved to evaluated the estimates, the same procedure is repeated 100 times, and a "Windows test" is computed too. In this test a random whole year is removed from the pretreatment sample to perform later evaluations; further details are in Section 9. From Table 2 LASSO-CV performs better[10].

TABLE 1	Random Points Test		Windows Test	
Square Errors	PSCM	LASSO-CV	PSCM	LASSO-CV
Average	0.5462	0.0585	1.4552	0.8912
Std	0.3739	0.0563	0.2079	0.4016

4.3 Spain

The Spanish law, Ley 35/2010, de 17 de September[11], had three main goals to declare in the same text: a) reduction of duality in the job market, fostering the creation of stable

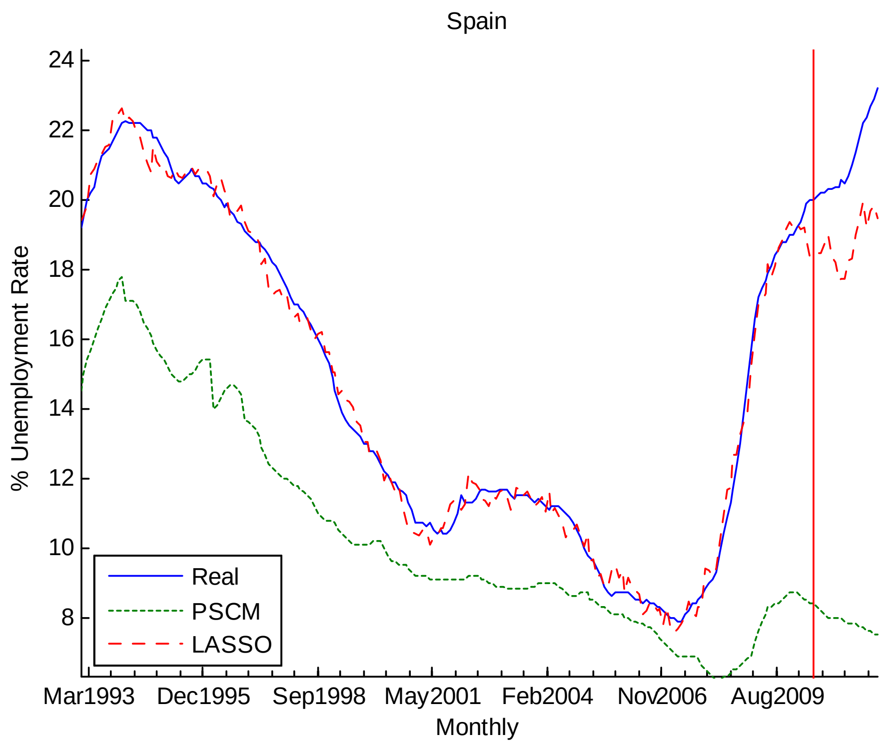
[10]Not suprise since PSCM is not passing the placebo test.

[11]Initially the reform was introduced by Real Decreto-ley 10/2010, de 16 de junio, with effectiveness on 18 of june.

and quality job; b) increased flexibility as an alternative for firing; and c) increased opportunities for the unemployed. However some disposals can be understood as a reduction in the firing cost, given that the cases of dismissal are augmented and their cost, in some cases, are reduced in a double manner; by the amount of money to pay for the company and because the government pay a proportion of the previously reduced amount[12].

Evidently there is an increase in unemployment for the Spanish case, probably around 2%. The coefficients are in Section 6 with details.

Figure 2



From Table 2 it is possible to conclude that LASSO-CV is doing better[13].

TABLE 2	Random Points Test		Windows Test	
Square Errors	PSCM	LASSO-CV	PSCM	LASSO-CV
Average	474.5739	3.9732	323.2603	6.4840
Std	147.1721	1.4171	397.1007	6.9255

Implementation of variation of the SCM are reported with not too much success, although some improvement, see Section 8.

[12] Excluding the so called "disciplinary dismissals".

[13] Not surprising since it is an example where PSCM is not working.

5 Conclusions

This paper's main contributions are: a) the inclusion of standard statistical learning methods in the tool set, in order to create alternative Counterfactual Series, providing theoretical support (in terms of pretreatment error minimization and prediction ability) and real support, via real case tests (it is another contribution of the paper, contribution d)); b) there are modificantions to SCM; and c) to the study of labor market changes, regarding the effects of a reduction of maximum standard workweek hours: Portugal (1996), and reductions in the labor cost of firing: Spain (2010).

The methodology, particularly based on LASSO, constructs a Counterfactual Series for a unit (i.e. a country, region or firm) by carefully selecting other units. The Counterfactual Series is compared with the real one, making it possible to appreciate the direction and magnitude of a change which affects a particular unit, such as a new law or a natural disaster. Its implementations are useful given that it is already in most of the statistical software packages.

For the empirical branch of the paper, it was found that unemployment decreased in Portugal by reducing the maximum standard workweek hours. In the Spanish case, reductions of cost firing increase the unemployment within a few months.

Further research is needed with the intention of increasing the inference in the counterfactual series, the improvements of the placebo tests, and the other test presented in this paper.

References

[1] Abadie, A., Diamond, A. and Hainmueller, J., 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California Tobacco Control Program," Journal of the American Statistical Association, American Statistical Association, vol. 105(490), pages 493-505

[2] Abadie, A., Diamond, A. and Hainmueller, J., (2014) "Comparative Politics and the Synthetic Control Method", American Journal of Political Science

[3] Abadie, A., and Gardeazabal, J., 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country," American Economic Review, American Economic Association, vol. 93(1), pages 113-132, March..

[4] Almer, C. and Winkler, R., 2012, "The Effect of Kyoto Emission Targets on Domestic CO2 Emissions: A Synthetic Control Approach." Working Paper. Universität Bern. Faculty of Business, Economics and Social Sciences. Department of Economics.

[5] Arlot, S. and Celisse, A., 2010, "A survey of cross-validation procedures for model selection." Statistics Surveys. Vol. 4 (20140) 40-79.

[6] Askenazy, P., 2013. "Working time regulation in France from 1996 to 2012," Cambridge Journal of Economics, Oxford University Press, vol. 37(2), pages 323-347.

[7] Bai, J. & Perron, P., 2003. "Computation and analysis of multiple structural change models," Journal of Applied Econometrics, John Wiley & Sons, Ltd., vol. 18(1), pages 1-22.

[8] Barone, G. & Mocetti, S., 2014. "Natural disasters, growth and institutions: A tale of two earthquakes", Journal of Urban Economics, Elsevier, vol. 84(C), pages 52-66.

[9] Billmeier, A., and Nannicini, T., 2013. "Assessing Economic Liberalization Episodes: A Synthetic Control Approach", The Review of Economics and Statistics, MIT Press, vol. 95(3), pages 983-1001, July.

[10] Cavallo, E., Galiani, S., Noy, I., and Pantano, J., 2013. "Catastrophic Natural Disasters and Economic Growth", The Review of Economics and Statistics, MIT Press, vol. 95(5), pages 1549-1561, December.

[11] Chow, G.C., 1960, "Tests of Equality Between Sets of Coefficients in Two Linear Regressions". Econometrica 28 (3): 591–605.

[12] Crimmann, A., Wießner, F. & Bellmann, L., 2010, "The German work-sharing scheme an instrument for the crisis." (Conditions of work and employment series, 25), Geneva, 46 S.

[13] Collier, D., 1993. "The Comparative Method." In Ada Finifter, ed., Political Science: State of the Discipline II, APSA.

[14] Doosan Encyclopedia (South Korea)

[15] Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R., 2004, "Least angle regression" (with discussion), Annals of Statistics 32(2): 407–499.

[16] Fisher, R. A., 1971, [1935]. "The Design of Experiments" (9th ed.). Macmillan.

[17] Freeman, R., 1998, "Work-sharing to full employment: Serious option or populist fallacy?" In: Freeman R, Gottschalk P (eds) "Generating jobs: How to increase demand for less-skilled workers", Russell Sage Foundation, New York, NY

[18] Hansen, Bruce E. & Racine, Jeffrey S., 2012. "Jackknife model averaging," Journal of Econometrics, Elsevier, vol. 167(1), pages 38-46.

[19] Hastie, T., Tibshirani, R., & Friedman, J., 2009, "The Elements of Statistical Learning. Data Mining, Inference, and Prediction." Springer. Second Edition.

[20] Hoerl, A. E. and Kennard, R.,1970, "Ridge regression: biased estimation for nonorthogonal problems", Technometrics 12: 55–67.

[21] Houpis, G., 1993, "The effect of lower hours of work on wages and employment". Discussion Paper No. 131, Centre for Economic Performance

[22] Hunt, J., 1996, "The response of wages and actual hours worked to the reductions of standard hours". NBER Working Paper 5716, Cambridge, MA

[23] Hunt, J., 1999, "Has Work-Sharing Worked In Germany?," The Quarterly Journal of Economics, MIT Press, vol. 114(1), pages 117-148, February.

[24] Ginsburg, T., 2012, "Legal Reform in Korea" (Routledge Advances in Korean Studies).

[25] Jacobson, T., and Ohlsson H., 2000. "Working time, employment, and work sharing: Evidence from Sweden," Empirical Economics, Springer, vol. 25(1), pages 169-187.

[26] Jones, R. H., 1980, "Maximum likelihood fitting of ARMA models to time series with missing observations", Technometrics, 22, 389-395.

[27] Kapteyn, A., Kalwij, A. & Zaidi, A., 2004, "The myth of worksharing," Labor Economics, Elsevier, vol. 11(3), pages 293-313, June.

[28] Kohavi, R., 1995, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection." Appears in the International Joint Conference on Artificial Intelligence (IJCAI).

[29] Houpis, G., 1993, "The efect of lower hours of work on wages and employment." Discussion Paper No. 131, Centre for Economic Performance

[30] Miller, A., 2002, "Subset Selection in Regression". Chapman and Hall/CRC.

[31] Lijphart, Arend (1971). "Comparative politics and the comparative method". American Political Science Review (The American Political Science Review, Vol. 65, No. 3) 65 (3): 682–693. doi:10.2307/1955513. JSTOR 1955513.

[32] Raposo, P. S. & van Ours, J. C., 2010. "How working time reduction affects jobs and wages," Economics Letters, Elsevier, vol. 106(1), pages 61-63, January.

[33] Rosenbaum, P.R. 2005. "Heterogeneity and Causality: Unit Heterogeneity and Design Sensitivity in Observational Studies." The American Statistician 59(2): 147-52.

[34] The American Statistician 59(2): 147-52.

[35] Rudolf, R., 2014. "Work Shorter, Be Happier? Longitudinal Evidence from the Korean Five-Day Working Policy," Journal of Happiness Studies, Springer, vol. 15(5), pages 1139-1163, October.

[36] Rudolf, R. & Seo-Young C., 2011. "The Gender-Specific Effect of Working Hours on Family Happiness in South Korea," Courant Research Centre: Poverty, Equity and Growth - Discussion Papers 77, Courant Research Centre PEG

[37] Sánchez, R., 2013. "Do reductions of standard hours affect employment transitions?: Evidence from Chile," labor Economics, Elsevier, vol. 20(C), pages 24-37

[38] Skuterud, M., 2007. "Identifying the Potential of Work-Sharing as a Job-Creation Strategy," Journal of Labor Economics, University of Chicago Press, vol. 25, pages 265-287.

[39] Tibshirani, R., 1996. "Regression shrinkage and selection via the lasso". Journal of the Royal Statistical Society, Series B 58 (1): 267–288. JSTOR 2346178.

Other References and Codes

- Spanish labor reform of 2010: Ley 35/2010, de 17 de September, de medidas urgentes para la reforma del mercado de trabajo. «BOE» núm. 227, de 18 de september de 2010, páginas 79278 a 79326 (49 págs.)
- Wikipedia, for general initial basic research.

Software and Hardware

Code is written in MATLAB 2014a. All codes I have created will be made available. The computer: Intel(R) Core(TM) i7-2600 CPU (3.40 GHz) with RAM 12 GB.

Appendix A: Data description

- Unemployment: Harmonized unemployment rate, monthly, total, all persons, seasonally adjusted. OCDE.

Countries in the sample Australia (AUSL), Austria (AUS), Belgium (BEL), Canada (CAN), Czech Republic (CR), Denmark (DEN), Finland (FIN), Germany (GER), Ireland (IRE), Italy (ITA), Japan (JAP), Luxembourg (LUX), Mexico (MEX), Netherlands (NET), Norway (NOR), Portugal (POR), Spain (SPA), Sweden (SWE), United Kingdom (UK) and United States (USA).

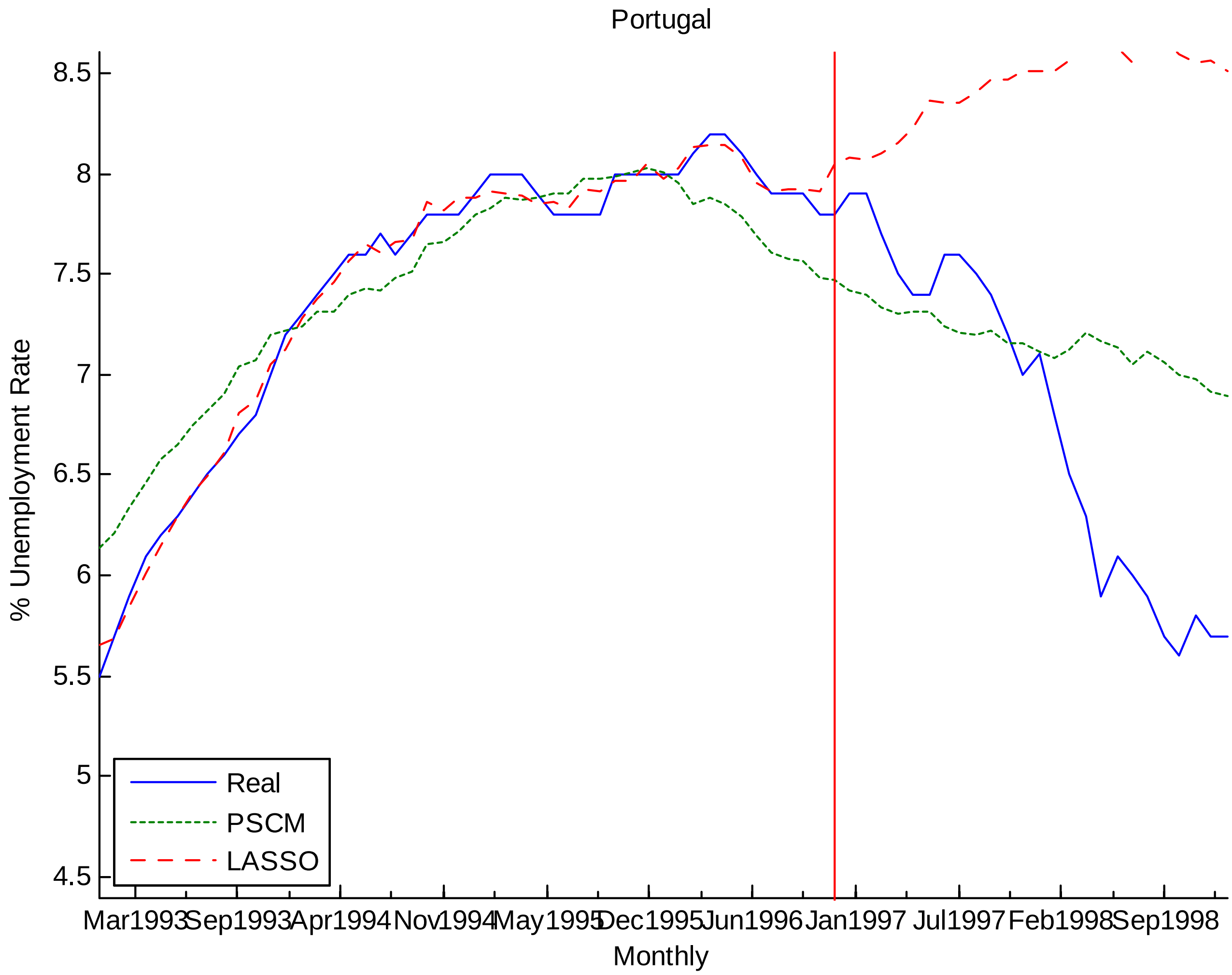
6 Appendix B: Table with coefficients

The further away t is from 1, the more different the models are (even if t=1, the models could be different), it is just an indicator. Although the usage of PSCM is not appropriate according to placebo tests results, see Section 7, I provide the coefficients an graphs.

Portugal

Country	AUSL	AUS	BEL	CAN	CR
PSCM	0	0	0.2822	0	0
LASSO-CV	-0.0768	-0.1788	0.7175	-0.1571	0.1089
	DEN	FIN	FRA	GER	IRE
PSCM	0	0	0	0	X
LASSO-CV	0.1820	0.0964	0.0944	-0.0630	X
	ITA	JAP	LUX	MEX	NET
PSCM	0.1613	0.2392	X	X	0.3173
LASSO-CV	-0.0410	0	X	X	0
	NOR	SPA	SWE	UK	USA
PSCM	0	0	0	0	0
LASSO-CV	0	0	0	0	0
	cte				
PSCM	0				
LASSO-CV	1.4372				
t	1.7159				

Figure 1 b



Spain

Country	AUSL	AUS	BEL	CAN	CR
PSCM	0	0	0	0	0
LASSO-CV	-0.0626	0.1317	-0.2175	-0.1962	0.4890
	DEN	FIN	FRA	GER	IRE
PSCM	0	1	0	0	X
LASSO-CV	0.2758	-0.5556	1.2944	-0.4751	X
	ITA	JAP	LUX	MEX	NET
PSCM	0	0	0	0	0
LASSO-CV	0.7008	-0.0434	0.1172	-0.2558	0.1458
	NOR	POR	SWE	UK	USA
PSCM	0	X	X	0	0
LASSO-CV	1.4156	X	X	0	0
	cte				
PSCM	0				
LASSO-CV	-7.4662				
t	6.3765				

7 Appendix C: Placebo tests

Placebo tests are common in the SCM literature, their implementation here is the same, however a new type is introduced, where only post-treatment square errors are reported; because LASSO-CV usually fits quite well, the pretreatment periods seems better to focus on the post-treatment periods.

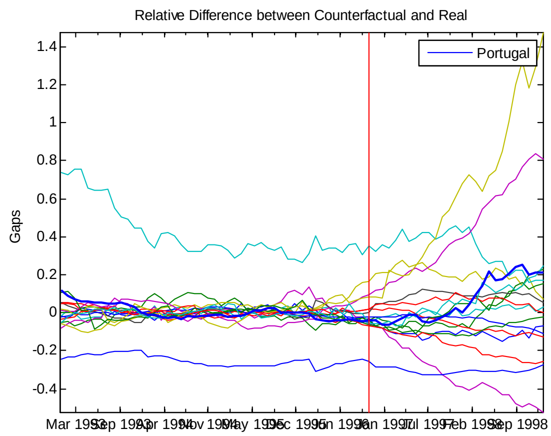
The results are portrayed by using of the gaps between real and counterfactual series, and histograms showing the ratio between post and pre-treatment mean square error, therefore, the further to the right is the studied unit the largest is supposed to be the change.

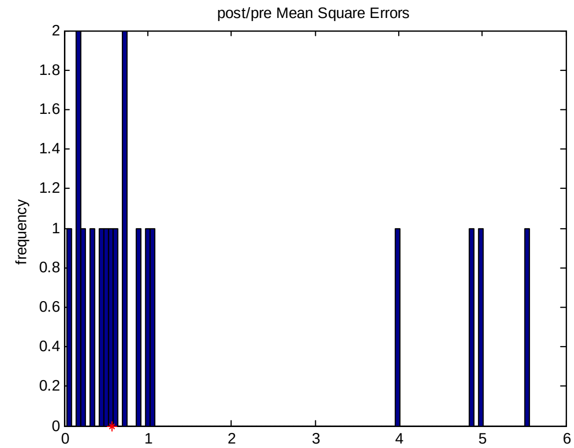
The units with larger errors than the treated unit are removed from the pool, to guarantee they do not affect the counterfactual series. In some cases it is impossible to claim that the treated unit is the one which suffers the largest shock, therefore it is not recommended use the methodology there. It is possible that by using PSCM a unit cannot pass the placebo test, but using LASSO-CV yes, here it is Portugal case.

The red star in the graphs points out the referred country positions. Portugal is not the country with the largest shock, in the Spanish case, it is the one with the largest shock.

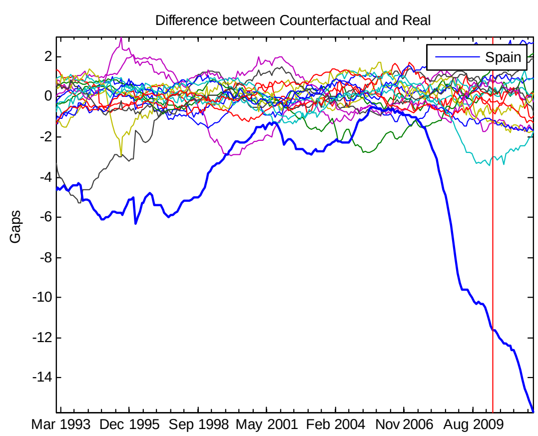
7.1 USING PSCM TO CALCULATE PLACEBO TEST

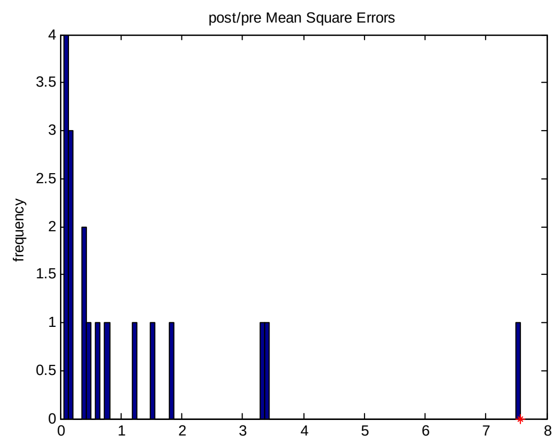
Portugal Portugal cannot be study with PSCM methodology.





Spain According to the placebo test, Spain could be treated, but according to the gaps in the graphs it is evident that this may not be a good idea.



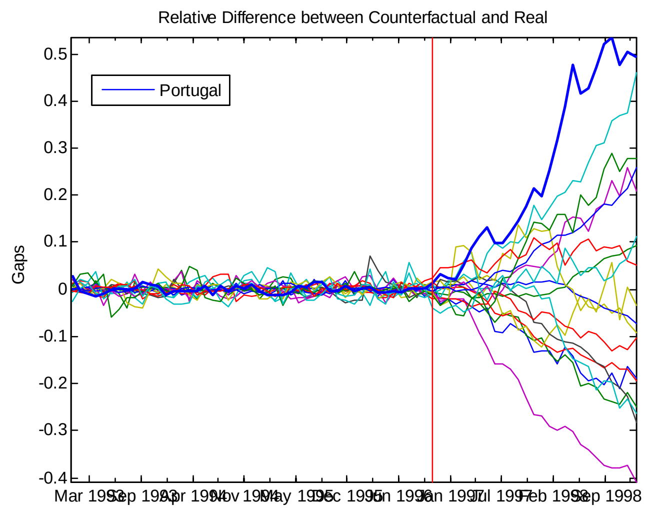


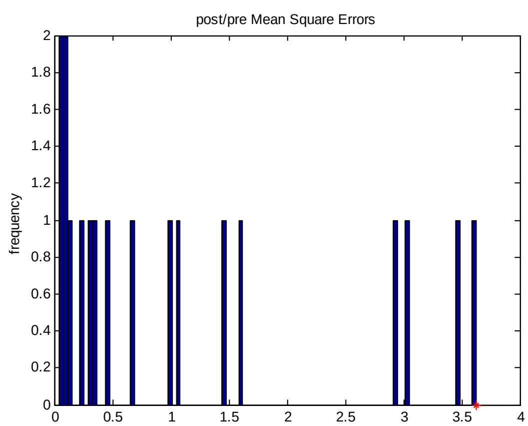
In other words, by using PSCM placebo test the conclusion is that PSCM should not be used for both Portugal and Spain.

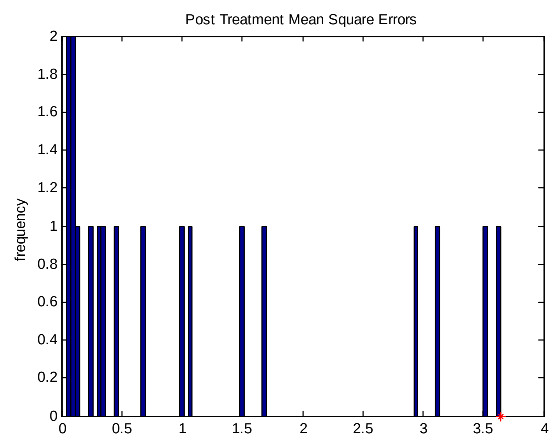
7.2 USING LASSO TO CALCULATE THE PLACEBO TEST

The units with largest errors are removed from the pool.

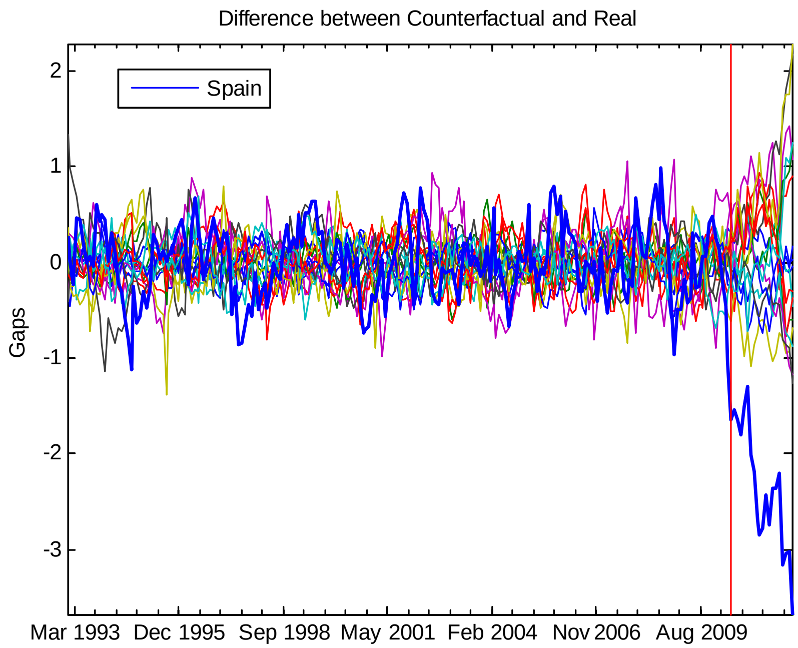
Portugal It is possible to appreciate, from all three next graphs, that Portugal is the country which has the largest change in the period.

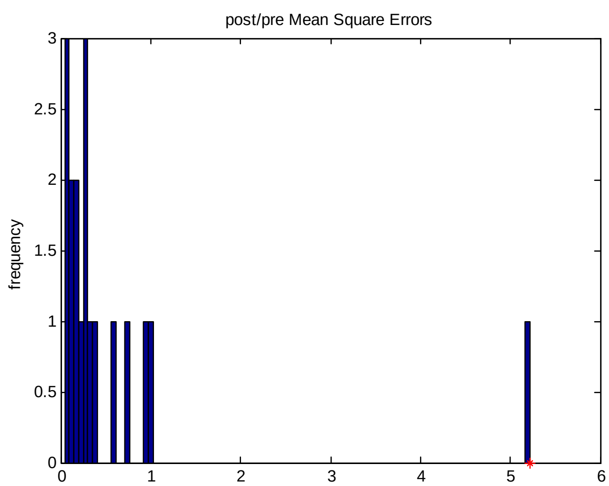


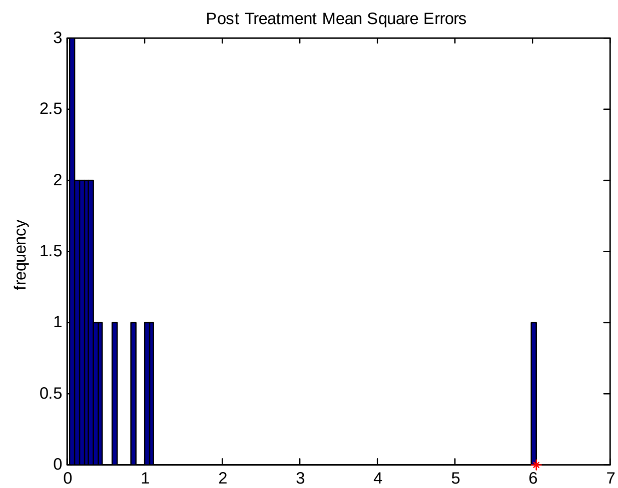




Spain It is possible to appreciate that Spain is the country which has the largest change in the period.



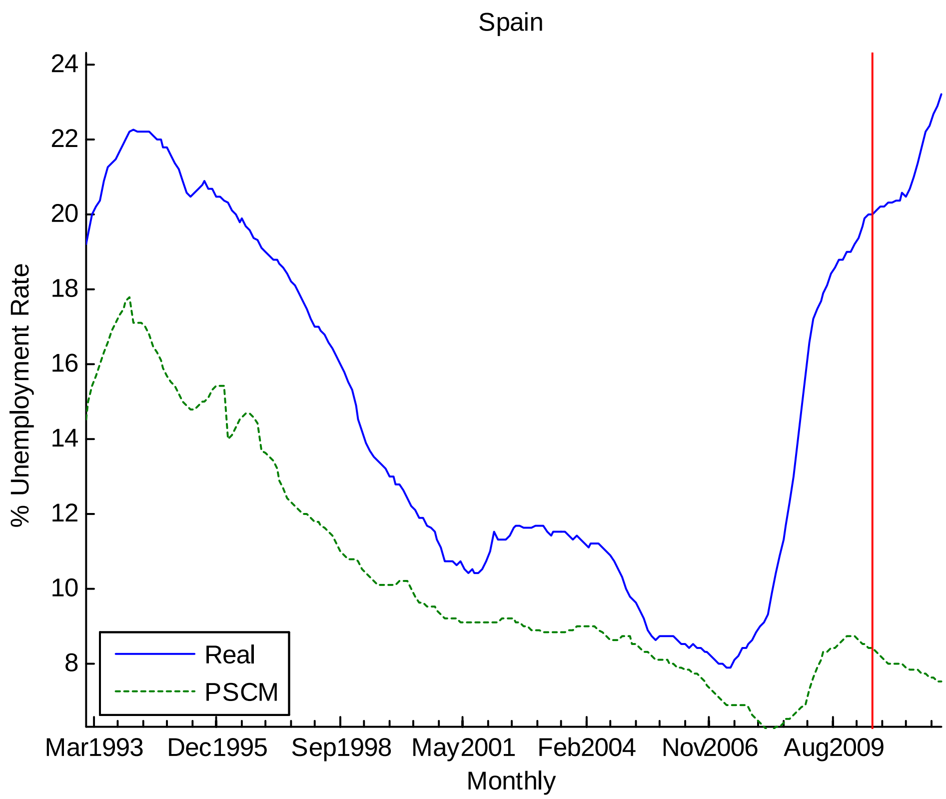




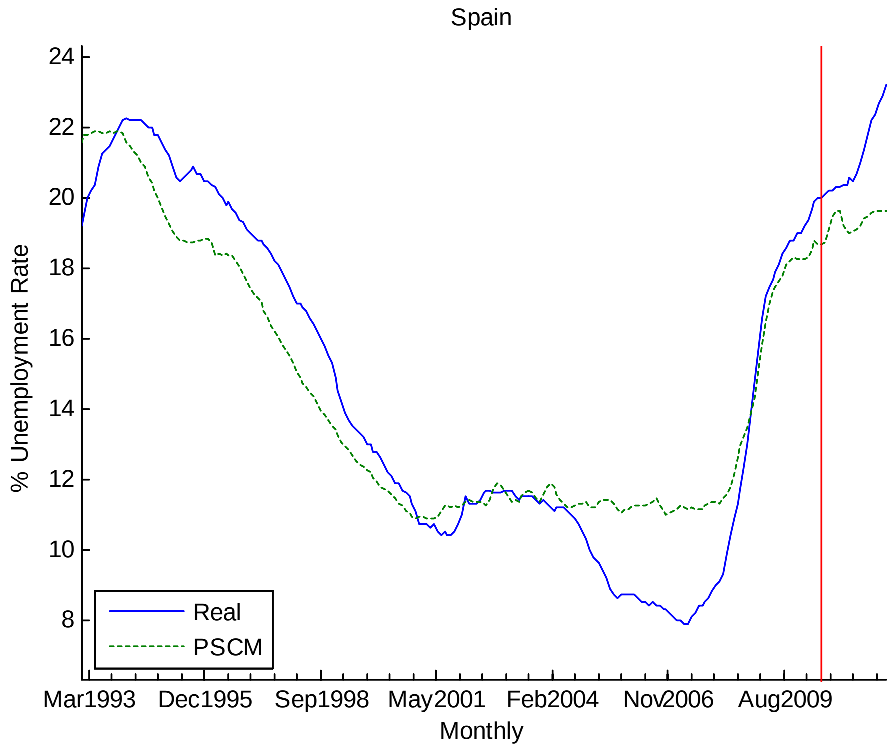
8 Appendix D: SCM with modifications

Here a case is depicted where PSCM is not working. I provide graphical evidence of two modifications with the aim to make possible the use of PSCM by adding a constant, it alleviate the problem. Secondarily CV is included to improve the coefficients obtained initially. The concrete algorithm is borrowed from Hansen and Racine 2012, the algorithm is know as "Jackknife model averaging", but no improvement is evident.

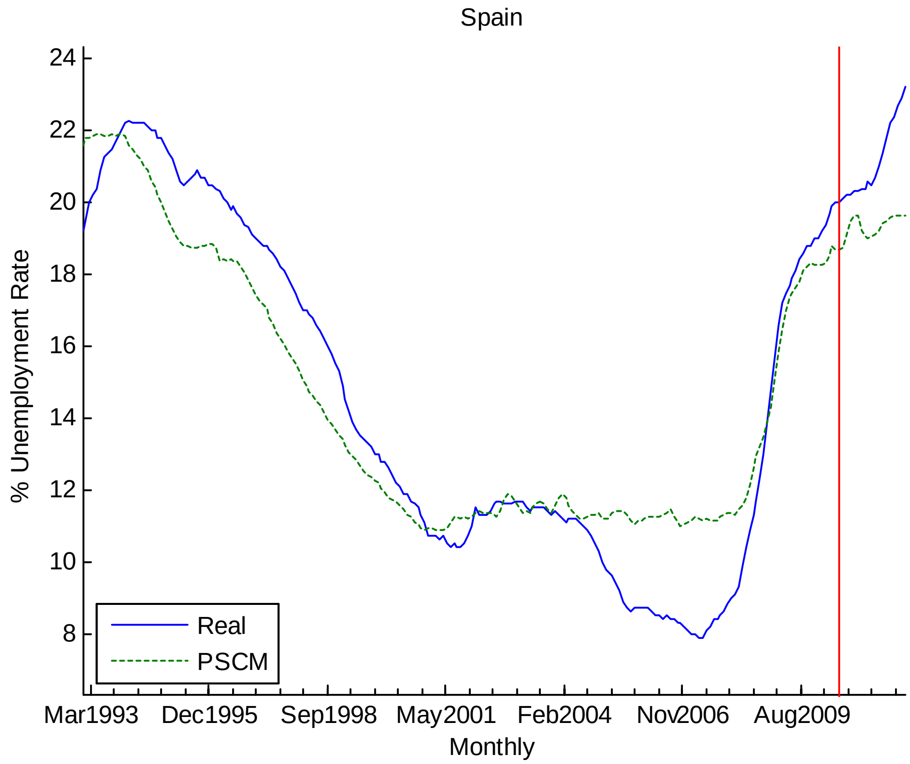
PSCM Initial case. Only Finland is used.



With a Constant term Spain is composed by a linear combination of Finland and Ireland.



With Cross-Validation Spain is composed by a linear combination of Finland and Ireland. In this case there is not a large difference in results. The methodology use is known as "Jackknife model averaging" by Hansen and Racine (2012).



9 Appendix E

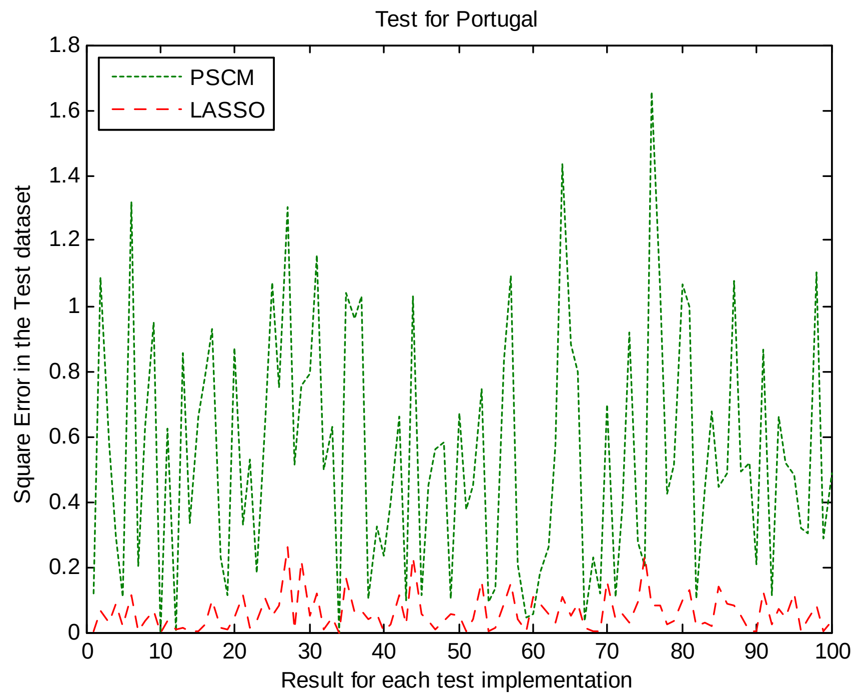
9.1 Random Points Test

The test presented here is created as follows:

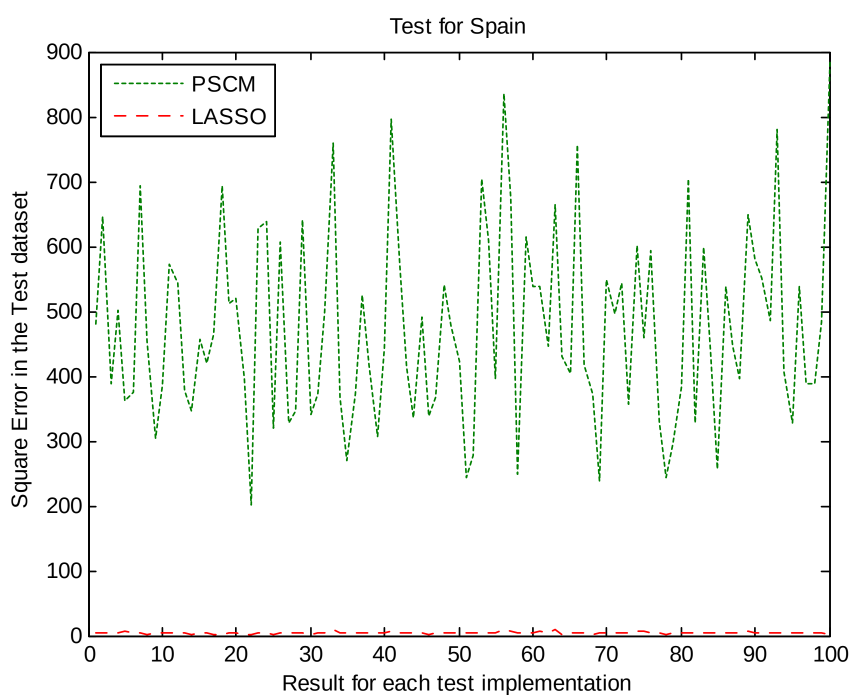
1. The pretreatment dataset is split in 10 randomly subsample.
2. One is reserved and not used for estimation. 9 subsamples are used for estimation.
3. The reserved data is used for assessing the previous estimate.
4. The results are gathered.
5. Previous steps are repeated 100 times.

Here the graphical results are shown for Portugal and Spain.

Portugal



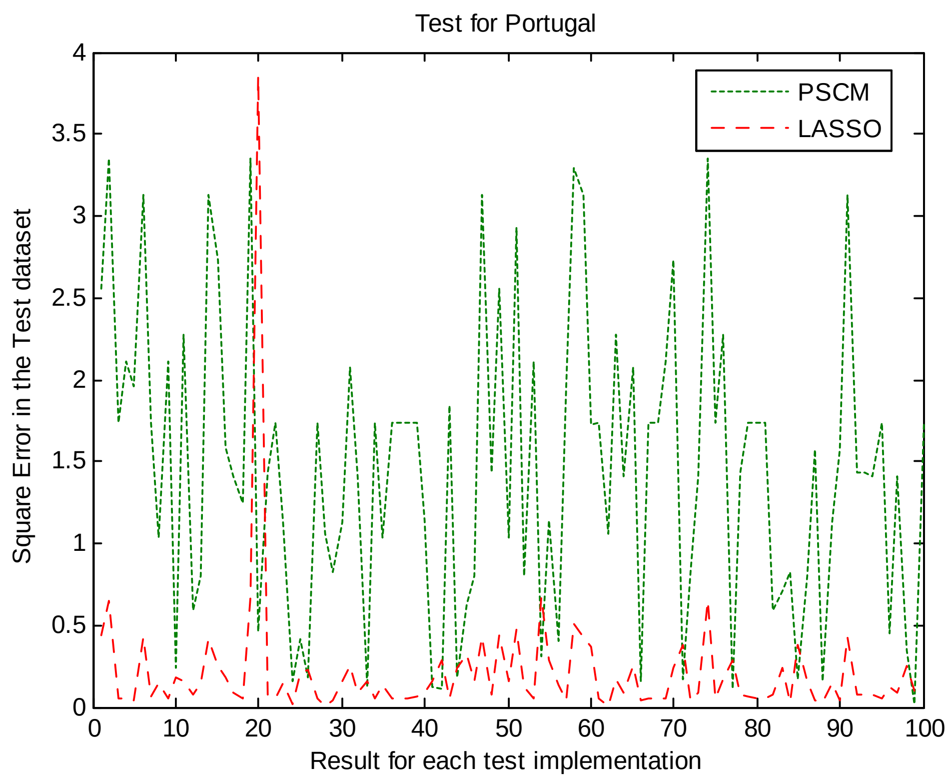
Spain



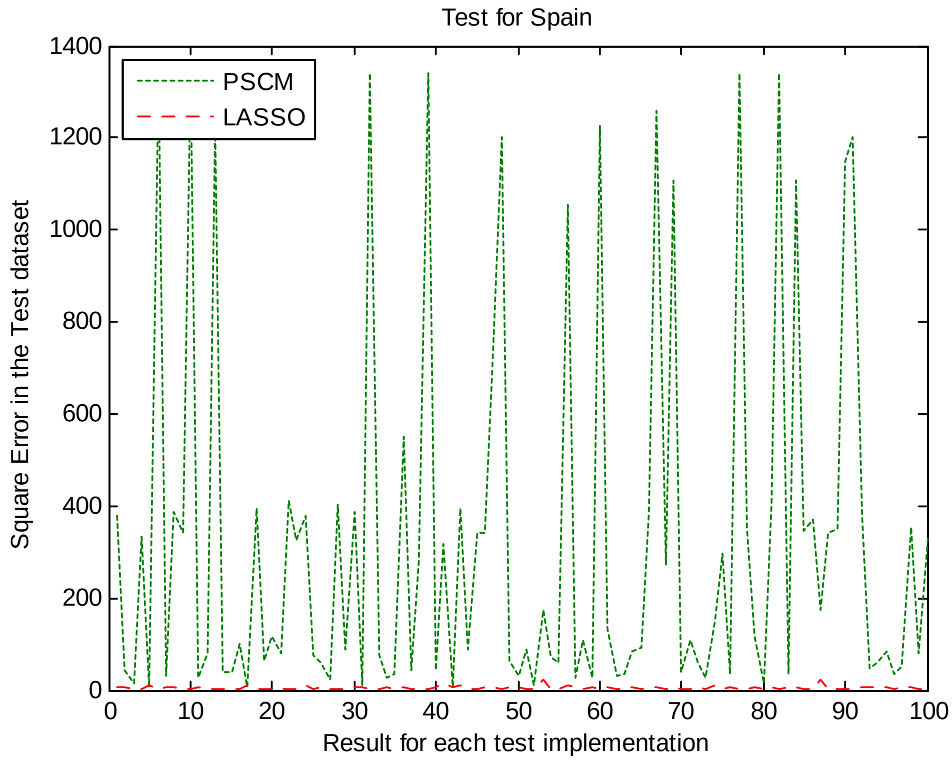
9.2 Windows Tests

What I refer to as a "Windows tests" consists of saving a set of data without been used for estimation, and the estimates are evaluated there. I do that 100 times. For Spain a random window of 12 months is taken each time. For Portugal, 6 months because of data availability. The results show that LASSO-CV performs quite a lot better in these cases than these cases that we already known are not the best to illustrate these tests.

Portugal

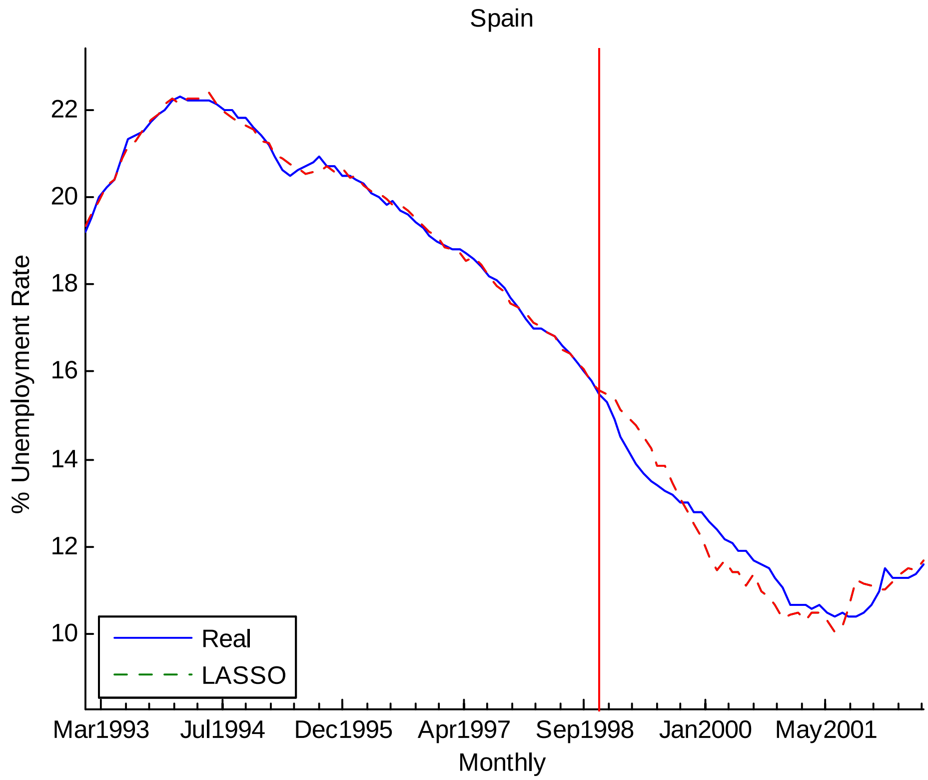


Spain



9.3 A Placebo Test in time

Here a placebo test is provided, where the breakpoint has been changed to one that doesn't not correspond to any known change. The result is that the counterfactual series is fairly close to the real one. Given that the original series has different breakpoints, a intermediate region has been selected.



10 Appendix H: The Bias–Variance trade-off and Model Selection

It is a well know problem in general. Here, standard notation is used. It can be briefly

summarized as follows: there are x_i points, with $i = 1, ...N$, they form the matrix X, for each of the previous values there is a y_i, they form the matrix Y. The data arises from the real but unknown model $Y = f(X) + \varepsilon$, there is a functional relationship $f(.)$, unknown, and a noise ε, with $E(\varepsilon) = 0$ and $var(\varepsilon) = \sigma^2$. The objective is to approach $f(.)$ by $\hat{f}(.)$. After some derivation the expected prediction error is:

$$E\left[\left(y - \hat{f}(x)\right)^2\right] = Bias[\hat{f}(x)]^2 + Var[\hat{f}(x)] + \sigma^2 \tag{8}$$

In expression 8 the Bias-Variance trade-off appears, the reduction of bias usually increases the variance and vice versa, both are usually affected by the complexity of the model. Therefore an equilibria should be found among the different models, i.e. Model Selection takes the best model. σ^2 is also known as irreducible error.

In other words, given a data set where a model is fitted, the corresponding Minimum Square Errors (MSE) on the data set are probably smaller than in the population or in another data set from the same source, i.e. with the same true functional relationship $f(.)$ and the other data set checks with the same estimation, i.e. $\hat{f}(x)$, obtained initially. A common problem in these situations is overfitting..

10.1 Adapted for our case

$\hat{f}(x, W, c) \equiv c + xW$

For all dataset we have, in LASSO-CV case:

$E\left[\left(Y_1 - \hat{f}(Y_0, \hat{W}^{LASSO}, c)\right)^2\right] = Bias[\hat{f}(Y_0, \hat{W}^{LASSO}, c)]^2 + Var[\hat{f}(Y_0, \hat{W}^{LASSO}, c)] + \sigma^2_{Y_1}$

$E\left[\left(Y_1 - c - Y_0\hat{W}^{LASSO}\right)^2\right] = Bias[c + Y_0\hat{W}^{LASSO}]^2 + Var[c + Y_0\hat{W}^{LASSO})] + \sigma^2_{Y_1}$

In the SCM or PSCM:

$E\left[\left(Y_1 - \hat{f}(Y_0, \hat{W}^{SCM}, 0)\right)^2\right] = Bias[\hat{f}(Y_0, \hat{W}^{SCM}, 0)]^2 + Var[\hat{f}(Y_0, \hat{W}^{SCM}, 0)] + \sigma^2_{Y_1}$

$E\left[\left(Y_1 - Y_0\hat{W}^{SCM}\right)^2\right] = Bias[Y_0\hat{W}^{SCM}]^2 + Var[Y_0\hat{W}^{SCM}] + \sigma^2_{Y_1}$