Use of Regression Discontinuity Design for the Estimation of the Impact of National Policies on the Development of Biomass Energy in China

Max-Sebastian Dovì

University of St Andrews, Queen's Gardens,St Andrews, KY16 9TQ, UK
dovi.msl@gmail.com

Regression Discontinuity Design (RDD) has become a popular econometric tool for evaluating impact and effectiveness of legislative measures**.** The main idea behind RDD is that discontinuities in the dependence of a set of data on a parameter across a cut-off value are basically due to the consequences of an event taking place at the cut-off. Thus, analyzing the time development of biomass energy in China, it is possible to estimate the impact of national policies by comparing differences across periods subject to different regulations. The relevant regression analysis introduces a dummy variable for each legislative bill considered: the corresponding coefficient in the ordinary least squares analysis provides the sought-after information on the effectiveness of the policy. Discontinuities in the rates of changes are also considered by introducing for each dummy variable two more terms representing a linear time evolution before and after the date of adoption of the new policy considered. Endogeneity problems are avoided by using the ratio of bioethanol production to corn and wheat production as a proxy variable for the overall biomass energy. Three milestone policies in the years 2002, 2007 and 2010 are considered. The estimated impacts and their statistical significance are evaluated using standard OLS techniques. The results show that RDD is indeed a powerful tool for an assessment of the effectiveness of new regulations in attaining their declared goals.

1. Introduction

In their seminal article, Thistlethwaite and Campbell (1960) introduced RDD to analyze the impact of merit awards (based on test scores) on the future academic career by comparing the academic outcomes of students with scores just below the cutoff with those just above it.

Analogously, it is possible to assess the consequences of a national policy by comparing the time evolution of suitable variables before and after the relevant regulations have been adopted (Lee and Lemieux, 2010). Differences in data that are close to but lie on either temporal side of the cut-off date can safely be assumed to depend only on the consequences of the introduction of the new policy.

Basically, the influence of an event (in this case the adoption of new regulations) can be assessed using a regression analysis that introduces a dummy variable for each legislative bill considered -once the regression has been carried out, the corresponding coefficient in the ordinary least squares analysis provides the sought-after information on the effectiveness of the event. Discontinuities in the rates of changes can also be considered by introducing for each dummy variable two more terms representing a linear time evolution before and after the date of adoption of the new policy considered. Indeed, the coefficients of the latter terms can be even more significant than the coefficients of the simple dummy variables because, due to the typical protractedness of bill approval procedures, the impact of a new policy can be distributed over an extended period of time.

Thus, the resulting regression discontinuity design for each new policy considered can be recast into the following regression equation:

$$r = a + \tau D + \beta_1(t-c) + \beta_2 D(t-c) + e \qquad (1)$$
$$c - h \le t \le c + h$$

A value of τ different from zero implies a discontinuity at the cut-off date, whereas a value of β_2 different from zero implies a change of slope, which switches from β_1 to $\beta_1 + \beta_2$.

Thus, estimating the coefficients α, τ, β_1, β_2 of the linear regression makes it possible to test the hypothesis of a shock effect at the cut-off point.

Ordinary least squares (OLS) can be used for the solution of this regression task provided the data points are limited to an interval (c-h)-(c+h), where c is the cut-off date considered. D is a binary dummy variable equal to zero if $t \le c$ and to one if $t \ge c$. The interval (bandwidth) h is the interval on either side of the cut-off point over which data are considered. It can be reduced provided the significance of the test (which can be derived from the p-values of the estimates) remains sufficient.

The value of τ resulting from the regression indicates the discontinuity estimated at the cut-off date, whereas the coefficients β_1 and β_2 provide the slopes on either side of the cut-off date. The coefficient α represents an uninfluential baseline value.

The application of RDD to China appears to be especially appropriate due to the particular structure of the Chinese decision-making process, the fiscal and normative regulatory framework and the government's influence on the internal financial market. Consequently, national policies, generally adopted after careful and years-long analyses, have a fundamental significance in the evolution of domestic production activities. Furthermore, the fact that the adoption of important decisions frequently coincides with the start of five-year-plans makes it possible not only to identify significant cutoff dates but also to consider adequate time intervals for verifying the impact of new norms and regulations.

2. The data used

2.1 Biomass energy data

The overall production of biomass energy in China (expressed as equivalent tons of coal) includes biogas, biomass pellets, and liquid biofuels. The traditional use of fuelwood, agricultural waste and animal dung for cooking, heating and lighting are not considered due to their limited responsiveness to governmental rulings and their declining importance.

Including all the three components in the regression analysis implies the necessity of considering a possibly large number of covariates. Indeed, factors such as the growth in the livestock sector and improved techniques for waste treatment must be considered when the time dependence of biogas is analysed. Similarly, the evolution of forest exploitation should be taken into account when estimating the time dependence of the biomass pellets production. If omitted, these covariates might give rise to serious endogeneity problems in the estimation of the coefficients α, τ, β_1, β_2 of Eq(1), because the error term ε would become time dependent and consequently correlated with the remaining covariates of the regression. On the other hand, there are no reliable data on these variables. As far as biofuels are considered, the production of biodiesel presents similar endogeneity problems Consequently, the production of bioethanol has been considered as a proxy for the overall biomass energy. Despite the increasing role of cassava in the production of bioethanol (Lauvena et al.. 2013), over 90 % of the bioethanol produced in China is made from wheat and corn (Anderson-Sprecher and Jiang Junyang, 2014). Therefore, the aggregate amount of wheat and corn harvests is the only additional explanatory variable that needs considering. Furthermore, with a view to avoiding the occurrence of another endogeneity factor (the reverse causality between the production of ethanol and the harvests of corn and wheat), the ratio of the production of bioethanol to the combined harvests of corn and wheat has been assumed as the response variable. Consequently,

$$y = \alpha + \tau D + \beta_1(t-c) + \beta_2 D(t-c) + \varepsilon \qquad (2)$$

where y is the ratio of the amount of bioethanol produced (r) to the overall harvest of corn and wheat (Q) in year t. The data for corn and wheat harvests are taken from the official website of the Chinese Statistical Yearbooks (2006, 2010, 2014), whereas the data for bioethanol production, also originally provided by the Chinese National Statistical Office, can be found in a report of the US Energy Information Administration. The relevant data are plotted in Figure 1.

The reliability of the data provided by the Chinese National Statistical Office has been occasionally questioned (Rawski, 2001). However, the quality of data improved considerably since 1998, when the very head of the NBS blamed the diffusion of deceptive intermediate reporting (jiabao fukuafeng, "wind of falsification and embellishment"), which was to influence the final official data very negatively. Furthermore, the selection of the

dependent variable, in addition to removing the reverse causality nexus, also tends to alleviate (and probably make irrelevant) the consequences of the inaccuracies contained in them. Indeed, as pointed out by Holz (2014), even if the official data were inaccurate, they are not statistically biased. In other words, inaccuracies are consistently recorded from one year to another.
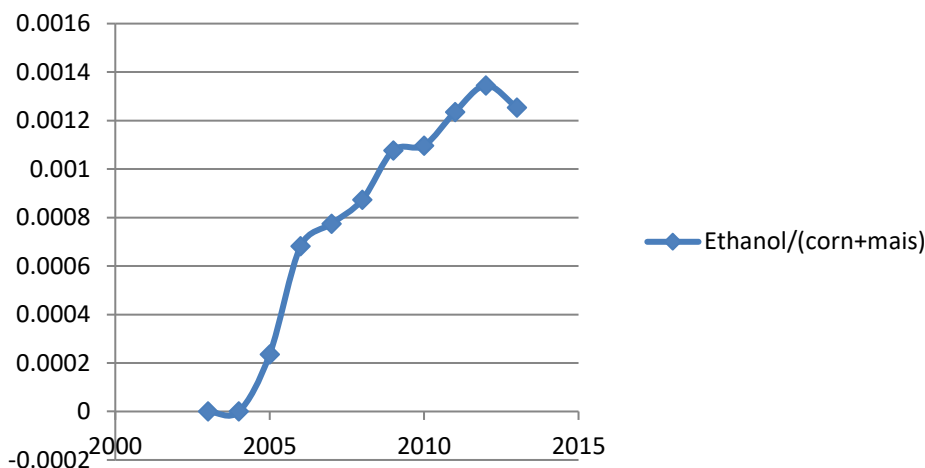
Ethanol/harvests of corn & wheat

Figure 1: Production of bioethanol in the years 2002-2013 (t bioethanol / t corn and wheat harvested)

Since the actual data set used is the ratio of bioethanol production to the overall output of the corn and wheat harvests, the dependent variable in the year *i* used for the regression is given by $y_i = \dfrac{r_i}{Q_i}$ where r_i and Q_i are the values of the production of bioethanol and of the harvests of corn and wheat in the year i reported by the Chinese National Statistics Service, if the values of r_i and Q_i are consistently over-reported by a fraction δ_1 and δ_2, i.e. $r_i^* = r_i - \Delta r_i = r_i - \delta_1 r_i = r_i(1 - \delta_1)$ and $Q_i^* = Q_i - \Delta Q_i = Q_i - \delta_2 Q_i = Q_i(1 - \delta_2)$ respectively, where r^* and Q^* indicate the true data, it follows that

$$y_i = \frac{r_i}{Q_i} = \frac{\dfrac{r_i^*}{(1-\delta_1)}}{\dfrac{Q_i^*}{(1-\delta_2)}} = \frac{r_i^*(1-\delta_2)}{Q_i^*(1-\delta_1)} \approx \frac{r_i^*(1-\delta_2)}{Q_i^*(1-\delta_1)} \approx \frac{r_i^*}{Q_i^*}(1+\delta_1)(1-\delta_2) = y_i^*(1+\delta_1)(1-\delta_2) \quad (3)$$

If the percent deviations are approximately equal (i.e. $\delta_1 \approx \delta_2 = \delta$, as assumed by Holz), the inaccuracies are small second-order deviations $y_i \approx y_i^*(1+\delta_1)(1-\delta_2) \approx y_i^*(1-\delta^2)$.

Thus, we can reasonably expect that any possible over-reporting by the Chinese Statistical Office can be safely neglected in the regression analysis examined.

2.2 Governmental acts considered
Three major governmental acts concern energy from biomass in China in the first decade of this century.
The so-called Renewable Energy Law of the People's Republic of China (Zhao and Yan, 2012) was issued in 2005 and can be regarded as the first law on renewable energy in China and as a direct response to the Kyoto Protocol, which entered into force on 16 February 2005. The law provides definitions of biofuels and commits China to the promotion of biomass fuels. In Chapter I the role of the State in the fostering of biomass energy is clearly underlined. The development of bioethanol and biodiesel is defined as a key priority for the government. In particular, a Renewable Energy Fund was established with a view to promoting "biofuel technology research and development, standards development and demonstration projects, support of biofuel investigation, assessment of raw materials resources, information dissemination and domestic related equipment manufacturing. (Report of the Global Subsidies Initiative, 2008)". It also specified the conditions

that biogas had to comply with to be fed-in into the public gas networks. Furthermore, biofuels were included in the National Renewable Energy Industry Development Guide Directory, which made it possible to obtain discounted loans and tax incentives for equipment and energy crops cultivation. Thus, the consumption tax and the value-added tax on bioethanol were either waived or refunded at the end of each year. Additionally, bioethanol plants were allowed to use "old grain" (grains reserved in national stocks no longer suitable for human consumption) and a minimum profit for each bioethanol plant was subsidized by the central government. These incentives are expected to have had a considerable impact on the development of biofuels in China.

In the years 2007 - 2008, a new set of policies were introduced and implemented by the NDRC (National Development and Reform Commission of the People's Republic of China, an agency under the Chinese State Council) and the Ministry of Finance. Two key issues were addressed: a revision of the subsidies policy and a limitation of biofuel production from edible crops, especially corn and wheat.

The revision of the subsidies policy implies flexible subsidies related to oil market price fluctuations, regulated by a Risk Fund in charge of preventing shocks from changes in gasoline market prices. The limitation in the use of edible crops grew out of the concern for food security. While the existing bioethanol plants were allowed to continue using wheat and corn as feedstocks, no additional plant should enter operation, nor were the existing plants allowed to increase their capacities. However, strong subsidies were foreseen for the production of bioethanol from non-cereal feedstocks originating from marginal lands, i.e. noncultivated land areas. Thus, while the two policies were intended to compensate each other, it was shown by Qiu et al. (2010) that China's bioethanol production target could be achieved only if "all potential non-cereal feedstocks and potential marginal lands would be used for bioethanol production". Indeed, this is an excellent opportunity to verify the efficacy of RDD in the evaluation of the impact of national policies on the development of economic activities.

The third governmental act considered is the Amendment Bill to the Renewable Energy Law issued in 2010. In addition to strengthening and consolidating the Renewable Energy Fund, established by the Renewable Energy Law of 2005, its main provisions concern an increased use of renewables within the overall electric power sector and a more detailed co-ordination of local-level development with national development plans. Furthermore, under the law, power companies were required to purchase all the renewable energy generated, even if in excess of the power demand on the grid. Economic penalties for companies failing to comply with this guaranteed-purchase requirement were also established.

It should be noted that in this case biofuels cannot be considered a proxy for the overall biomass energy. Indeed, apart for the strengthened Renewable Energy Fund, biofuels should be regarded as competitors with other biomass energy carriers, which are more suitable for power generation. Thus, this governmental act has been considered to verify the correctness of the RDD analysis and thus to further validate this approach.

3. Results

The dates of adoption of each of the three governmental acts considered in the previous section serve as cut-off values of the corresponding RDD analysis. The bandwidth includes 6 data points for each cut-off date analysed. In the tables below are reported the point values of the coefficients of the covariates, as well as the confidence intervals and the corresponding p-values, which provide a measure of the statistical significance of the estimates obtained. Following general usage, estimates are considered statistically significant if their p-value < 0.05 and statistically highly significant if the p-value < 0.01, marked with one and two asterisks respectively.

The analysis of the three cases is carried out separately for each of them. Standard errors and p-values for coefficients whose magnitudes are less than 0.0001 are not reported.

3.1 Impact of the Renewable Energy Law

Table 1: Coefficients of the regression equation

	Coefficients	Standard Error	p-value	Lower 95 %	Upper 95 %
α (Intercept)	-6×10^{-19}	---	---	---	---
τ (Covariate #1)	0.000341	0.0005	0.565472	-0.00181	0.00249
β_1 (Covariate #2)	-2.3×10^{-19}	---	---	---	---
β_2 (Covariate #3)	0.000201	0.0002	0.421956	-0.00066	0.001062

While the intercept and the initial slope are correctly identified, the high correlation between τ and β_2 makes it impossible to estimate both in a statistically significant manner. Indeed, the correlation factor is -0,964, which

explains the large p-values and confidence intervals evaluated. However, since both coefficients imply a strong impact of the governmental act, there is no evident contradiction. The presence of an over-parameterization can be checked by eliminating the dummy variable D as an explanatory variable in the regression analysis. Indeed, the relevant regression provides the following estimates:

Table 2: Coefficients of the simplified regression equation

	Coefficients	Standard Error	p-value	Lower 95 %	Upper 95 %
α (Intercept)	1.2×10^{-5}	---	---	---	---
β_1 (Covariate #2)	-7.4×10^{-5}	---	---	---	---
β_2 (Covariate #3)	0.000178	3.61×10^{-5}	0.0079**	7.8×10^{-5}	0.000279

The estimation of both intercept and slope is now statistically significant. Indeed, the estimate of β_2 (which is of greater importance for evaluating the impact of the Renewable Energy Law) is highly significant being lower than 0.05. Both intercepts and initial slope are close to zero, which corresponds to a negligible level of bioethanol production.

3.2 Impact of the supporting policies of 2007

Table 3: Coefficients of the regression equation

	Coefficients	Standard Error	p-value	Lower 95 %	Upper 95 %
α (Intercept)	0.00138	2.45×10^{-5}	0.01130*	0.001069	0.001692
τ (Covariate #1)	-0.00029	2.58×10^{-5}	0.057201	-0.00061	4.14×10^{-5}
β_1 (Covariate #2)	0.000203	1.19×10^{-5}	0.03733*	5.14×10^{-5}	0.000354
β_2 (Covariate #3)	-0.0001	1.28×10^{-5}	0.078126	-0.00027	5.88×10^{-5}

The results seem to convey that of the two effects generated by the supporting policies adopted in 2007, the one with negative consequences on the production of bioethanol (i.e. the limitation on the production of bioethanol from edible crops) seems to prevail in the short term. Indeed, while the rate of growth (as given by the sum of β_1 and β_2) is hardly reduced (less than 0.5 %), there is a substantial production drop at the cut-off date. In other words, the growth in the production of bioethanol (possibly determined by general market conditions) is only temporarily affected by the new norms, the diversification of feedstocks being favoured by the additional provisions contained in the law. All estimates are either statistically significant (p-value lower than 0.05) or close to the significance limit.

3.3 Impact of the Amendment to the Renewable Energy Law

Table 4: Coefficients of the regression equation

	Coefficients	Standard Error	p-value	Lower 95 %	Upper 95 %
α (Intercept)	0.00144	4.81921×10^{-6}	0.0021**	0.00138	0.00150
τ (Covariate #1)	-9.77×10^{-5}	---	---	---	---
β_1 (Covariate #2)	0.00014	2.33766×10^{-6}	0.0107*	0.00010	0.00017
β_2 (Covariate #3)	-0.00020	2.51545×10^{-6}	0.0079**	-0.00023	-0.00017

In this case, the coefficient of the dummy variable D is vanishingly small, so that there is no apparent discontinuity at the cut-off date. However, there is a considerable, statistically significant (or highly significant) change of slope (p-value of β_1, < 0,05 and p-value of β_2, < 0,01) across the cut-off date. Furthermore, the negative value of $\beta_1 + \beta_2$ corresponds to the expected effect of a reduction of bioethanol production due to the competitive use of biomass for power generation. Thus, the overall impact of the amendment to the Renewable Energy Law, which is expected to have a negative influence on bioethanol production, is well reflected by the values of the coefficients of the regression.

4. Conclusions

The application of Regression Discontinuity Design to the evaluation of the impact of recent governmental acts on the production of bioethanol in China has been shown to provide a powerful tool for the quantitative assessment of the consequences of legislative decisions well beyond the information that a simple visual inspection can provide.

However, unlike the original formulation, RDD, when applied to bills passed by competent authorities, strictly needs changes in slope being taken into consideration. Indeed, as shown in all the cases considered, the market response to bills under consideration by a legislature can well anticipate the expected approval. In other words, the consequences of an impending bill can be distributed in an approximately uniform way over the time interval it requires to become law.

While the introduction of slope changes across the cut-off date greatly enhances the assessment capacity of RDD, it also poses the issue of how to combine discontinuities and slope changes for an overall impact assessment when both are present. This is all the more important if a strong correlation exists between the two terms. Indeed, since the two terms are largely interchangeable, the resulting index must be independent of any combination of them.

Thus, the development of a generally agreed methodology on this issue is necessary before Regression Discontinuity Design becomes a recognized assessment tools of governmental acts.

Acknowledgments

References

Anderson-Sprecher A., Jiang J., 2014, GAIN Report CH14038, USDA Foreign Agricultural Service, Global Agricultural Information Network, <gain.fas.usda.gov/Recent%20GAIN%20Publications/Biofuels%20Annual_Beijing_China%20-%20Peoples%20Republic%20of_11-4-2014.pdf>, Accessed 28/09/2016.

Cai Y., 2000, Between State and Peasant: Local Cadres and Statistical Reporting in Rural China, The China Quarterly, 163, 783-805.

Holz C.A., 2014, The quality of China's GDP statistics, China Economic Review, 30, 309–338.

Qiu H., Huang J., Yang J., Rozelle S., Zhang Y., Zhang Y., Zhang Y., 2010, Bioethanol development in China and the potential impacts on its agricultural economy, Applied Energy, 87, 76–83.

Lauvena L.P., Beibei L., Geldermann J., 2013, Investigation of the Influence of Plant Capacity on the Economic and Ecological Performance of Cassava-based Bioethanol, Chem. Eng. Trans., 35, 595-600.

Lee D.S., Lemieux T., 2010, Regression Discontinuity Designs in Economics, Journal of Economic Literature, 48, 281-355.

Martinot E., Li Junfeng, 2010, China's Latest Leap: An update on renewables policy, Renewable Energy World, <www.renewableenergyworld.com/articles/print/volume-13/issue-4/solar-energy/renewable-energy-policy-update-for-china.html>, Accessed 04/03/2016.

Thistlethwaite D.L., Campbell D.T., 1960, Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment, Journal of Educational Psychology, 51, 309–317.

Rawski T.G., 2001, What is happening to China's GDP statistics?, China Economic Review, 12, 347–354.

Report of the Global Subsidies Initiative, 2008, Biofuels – At What Cost? Government support for ethanol and biodiesel in China, International Institute for Sustainable Development, ISBN: 978-1-894784-24-5.

U.S. Energy Information Administration, 2014, International Energy Statistics <www.eia.gov/cfapps/ipdbproject/iedindex3.cfm?tid=79&pid=80&aid=1&cid=regions&syid=2000&eyid=2012&unit=TBPD>, Accessed 04/03/2016.

National Bureau of Statistics of China, 2006, China Statistical Yearbook <www.stats.gov.cn/tjsj/ndsj/2006/indexeh.htm>, Accessed 04/03/2016.

National Bureau of Statistics of China, 2010, China Statistical Yearbook <www.stats.gov.cn/tjsj/ndsj/2010/indexeh.htm>, Accessed 04/03/2016.

National Bureau of Statistics of China, 2014, China Statistical Yearbook <www.stats.gov.cn/tjsj/ndsj/2014/indexeh.htm>, Accessed 04/03/2016.

Zhao Z.-Y., Yan H., 2012, Assessment of the biomass power generation industry in China, Renewable Energy, 37, 53-60.