Response letter to the 2nd Referee Report for the article entitled as "Gendered Economic Policy Making: The Case of Public Expenditures on Family Allowances"

I am thankful for the useful comments and feedback received from the second anonymous referee. The comments are indeed useful to improve my paper's quality. Please see the explanations below for the issues raised by the referee. The referee's comments are provided in bold characters.

1.The use of an arbitrarily chosen critical mass threshold (15 %, 20 %, 25 % and 30 %) that is not based on a formal test is unsatisfactory. You may consider performing a structural break analysis with an unknown threshold level to formally test for the critical mass argument. With respect to organization, it is advisable to consider including the structural break analysis as a separate section prior to the presentation of the baseline results, which would strengthen your main argument on the relationship between the critical mass of female MPs and the expenditure on family allowance.

Considering a structural break in a time series framework analysis, examining one country can be straightforward. However, in the case of my paper where a panel dataset is explored, it might be difficult for all countries to "move" or "jump" at the same time since countries are characterized by different political structure and there is an obvious heterogeneity (It was the reason why fixed effects models are also estimated in the paper to control cross-country heterogeneity). If the panel data was at individual level within a country, a jump could be again possible since individuals would behave similarly as a result of a political change. However, this is again not the case in the paper where different countries are examined using a country level panel dataset. Second problem is that the date of the structural break may not known beforehand. Although these mentioned problems might occur, there might be three different solutions, however, to perform a structural break analysis with an unknown threshold level to formally test for the critical mass argument. The first solution proposed is to estimate for possible structural break with unknown date in each country separately as an examination in a time series framework.

1.1 First solution: *Structural break for every country*

Initially, Andrews (1993) proposed the following test. Suppose that;

$y_t = x_t'\beta_1 + \varepsilon_t \ , \ t = 1, \ldots\ldots, \tau$ (1)

$y_t = x_t'\beta_2 + \varepsilon_t \ , \ t = \tau + 1, \ldots\ldots, T$ (2)

β_1, β_2, and x_t are at $k \times 1$ dimension. There is a single breakpoint, which is τ. Assume the x's are stationary and weakly exogenous and the ε's are serially uncorrelated and homoscedastic. Then we examine the following hypothesis;

$H_0: \beta_1 = \beta_2$

If τ is known the F-statistic is:

$F_T(\tau) = (T-2k)\,[SSR_{1:T} - (SSR_{1:\tau} + SSR_{\tau+1:T})] / (SSR_{1:\tau} + SSR_{\tau+1:T})$ (3)

, where T is the number of years, k the number of regressors, $SSR_{1:T}$ is the sum of squared residuals of regression (1) and similarly $SSR_{\tau+1:T}$ is the sum of squared residuals of regression (2). The F-statistic follows asymptotically $\chi^2(k)$ under H_0. In the case where τ is unknown, Quandt (1960) showed that the likelihood ratio statistic corresponding to $H_0: \beta_1 = \beta_2$ is:

$$QLR_T = \max_{\tau \varepsilon \{\tau_{min}, \ldots, \tau_{max}\}} F_T(\tau)$$

(4)

Andrews (1993) showed that under appropriate regularity conditions, the QLR statistic, also referred to as a $SupLR$ statistic, has a nonstandard limiting distribution and specifically under H_0 is:

$$QLR_T \xrightarrow[D]{} \frac{sup}{r \varepsilon [r_{min}, r_{max}]} \left(\frac{B_k(r)' B_k(r)}{r(1-r)}\right)$$

(5)

,where $0 < r_{min} < r_{max} < 1$ and B_k (.) is a "Brownian Bridge" process defined on [0,1] and sup is the least upper bound of a set S.

Accordingly In Table I, the QLR test statistic, along with the p-values, is reported[1]. It becomes obvious that every country is exposed in different year of structural break. Then the following regression for various thresholds is estimated.

Table I. Structural Breaks For Every Country Using base Sample

Country	Break Year	QLR test statistic
Australia	1991	67.5117
		(0.0000)
Belgium	1985	225.941
		(0.000)
Canada	1985	583.5519
		(0.000)
Denmark	1987	49.4299
		(0.000)
Finland	1991	75.5060
		(0.000)
France	1986	353.4097
		(0.000)
Greece	1990	314.7958
		(0.0000)
Ireland	1992	63.929
		(0.000)
Italy	1986	115.593
		(0.0000)
Japan	2001	220.1292
		(0.0000)
Luxemburg	1992	117.5735
		(0.000)
New Zealand	1996	171.925
		(0.0000)
Netherlands	1987	62.3748
		(0.0000)
Portugal	1990	83.8733
		(0.0000)
Spain	1986	92.0637
		(0.0000)
Sweden	1991	52.212
		(0.0000)
Switzerland	1996	122.7585
		(0.0000)
United Kingdom	1991	197.0192
		(0.0000)
USA	1992	81.1321
		(0.0000)

Not: p-values are represented within brackets

[1] Due the longest time dimension compared to other two samples, structural breaks using the first sample of the paper, which is covered from 1980 to 2008, are shown here. Other structural break tables can be obtained using other samples as well.

$$y_{it} = a_i + \beta thresh_{it} + \gamma' \sum_{j=1}^{J} D_{jit} + \delta' \sum_{j=1}^{J} D_{jit} \times thresh_{it} + \phi' X_{it} + country + year \quad (6)$$

,where y_{it} is the family allowances, *thresh* indicates the threshold level of the female parliamentarian percentage i.e. 20, 30 per cent, D_{jit} is the dummy for country *i* and time *t*, taking value 1 for the post-period (after the structural break) and 0 for the pre-period for j=1…J countries. Thus, the regressions control for the dummy $D_{jit,}$ and the interaction term of $D_{jit} \times thresh_{it}$ for all countries. In Table II, the coefficient of the threshold (β) for every threshold as well as the F-statistic for the interaction term $D_{jit} \times thresh_{it}$ are reported. In all cases the null hypothesis of the non-significance of the joint interaction term $D_{jit} \times thresh_{it}$ for all countries is rejected, confirming that the year of the structural breaks found in Table I exist. On the other hand, the threshold coefficient is significant for a percentage up to 28, while it becomes significant for 29 per cent and higher which is near to the final finding (30 per cent) of the paper.

Table II. Threshold and Structural Break Tests.

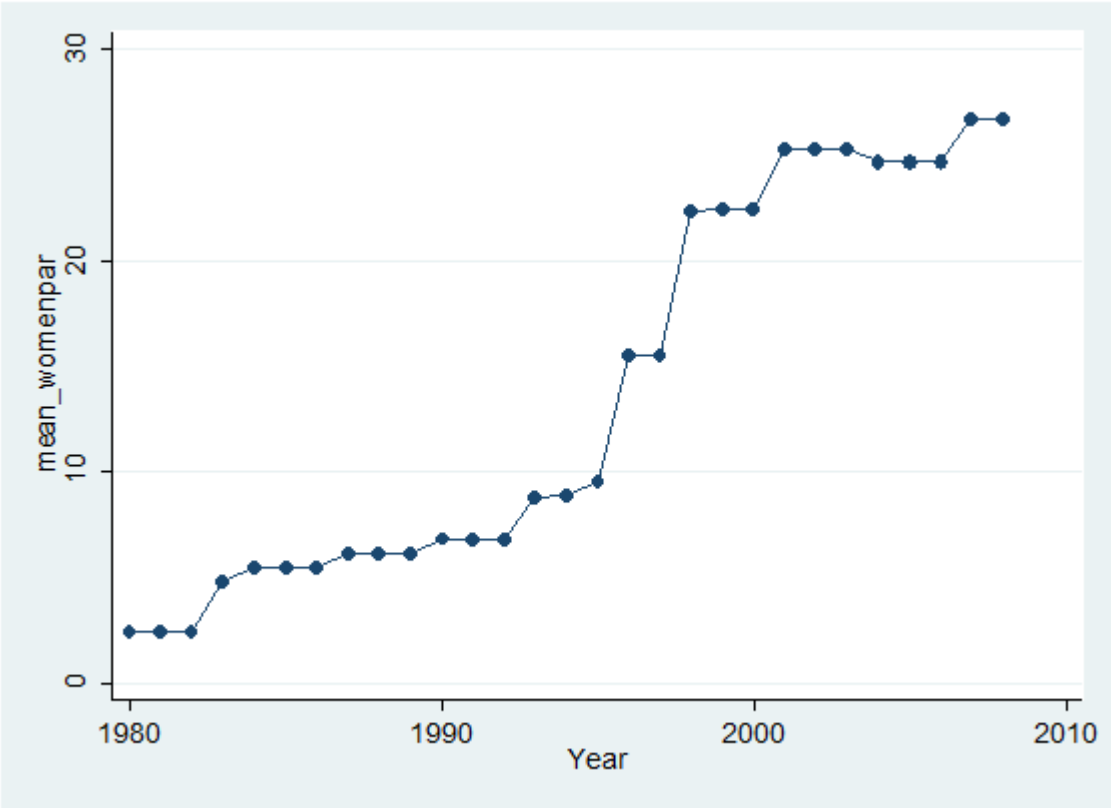
Threshold (%)	Coefficient β.	F- statistic for the joint interaction term $D_{jit} \times thresh_{it}$
20	-0.1060	4,507.86
	(0.0622)	[0.000]
25	0.0028	23,642.52
	(0.0672)	[0.000]
26	-0.02765	7,740.26
	(0.0807)	[0.000]
27	0.0950	3,362.23
	(0.0997)	[0.000]
28	0.1624	2,329.60
	(0.1032)	[0.000]
29	0.2452**	24,117.66
	(0.1013)	[0.000]
30	0.3003***	67,950.03
	(0.0598)	[0.000]
31	0.3288***	1,634.00
	(0.0567)	[0.000]
32	0.3070***	4,303.36
	(0.0745)	[0.000]
33	0.3173***	4,114.25
	(0.0746)	[0.000]

Robust standard errors within brackets, p-values within square brackets, *** and ** indicate significance at 1% and 5% level.

1.2 Second solution: *Total Countries Graph Approach*

The second solution is to obtain the mean of the women participation rates over the countries for every year and present into a plot. In Figure 1, it can be seen that a possible structural break takes place in the year of 1995. However, finding the year of the structural break that takes place does not imply the fact that also a specific threshold is significant too. For instance, we have to examine whether the threshold of i.e. 20 per cent becomes also significant after 1995 or not.

Figure 1. Average women participation for all countries



Based on the Figure 1 the following regression is estimated.

$$y_{it} = a_i + \beta_1 thres_{it} + \beta_2 D_{it} + \gamma D_{it} \times thres_{it} + \phi' \mathrm{X}_{it} + country + year \quad (7)$$

, where the variables are defined a in (6), while only one dummy is considered for all countries that takes value 1 for the post-period, after the structural break takes place in 1995 and takes value 0 for the pre-period. In all cases coefficient β_2 is statistically significant and positive confirming that a structural break takes place after 1995 and indicating that after this break a significant increase on family allowances is observed.

Regarding coefficient $\beta_{1,}$ it becomes positive and significant when the women participation passes the 26 per cent and after. Previously it has been found a threshold equal or higher than 29 per cent is significant. However, since the political changes happen in different time periods for each country, the first solution that is based on structural breaks in different years for every country is more realistic and reliable to my opinion.

Table III. Estimation of (7) For Various Threshold and Structural Break in 1995

Threshold %	Coefficient β_1	Coefficient β_2
20	0.0284	0.4360**
	(0.1051)	(0.1644)
25	0.1472	0.4323**
	(0.0973)	(0.1705)
26	0.1599	0.4470**
	(0.0961)	(0.1747)
27	0.2316**	0.4106**
	(0.0876)	(0.1716)
28	0.2515***	0.4009**
	(0.0797)	(0.1768)
29	0.2871***	0.3899**
	(0.0769)	(0.1748)
30	0.2839***	0.3971**
	(0.0728)	(0.1750)
31	0.2932***	0.3857**
	(0.0714)	(0.1682)
32	0.3683***	0.3275**
	(0.0723)	(0.1370)
33	0.3061***	0.3777**
	(0.0837)	(0.1492)
34	0.1361*	0.0306
	(0.0691)	(0.0214)
35	0.1373*	0.0302
	(0.0675)	(0.0211)

Robust standard errors within brackets, ***, ** and ** indicate significance at 1%, 5% and 10% level.

1.3 Third solution: Kernel-weighted linear local polynomial smoothing regression

In the case where a non-randomized assignment is not applicable, a linear piecewise spline regression similar to a regression discontinuity (RD) is applied in this study in order to identify the level of the threshold of women participation which causally affects the family allowances. Regression discontinuity design (RDD) has grown increasingly common in economics and has been applied to identify causal effects in a wide variety of contexts where traditional approaches have difficulty (Caughey and Sekhon, 2011 for review).

The RD design was first introduced by Thistlethwaite and Campbell (1960) in their study of the impact of merit awards on future academic outcomes and since then it has been applied in various studies (Imbens and Lemieux, 2008; Lee and Lemieux). The reason why RD is followed is to examine and identify the causal effect of the women participation on family allowances. More specifically, values just above the threshold and just below the threshold are compared controlling for the same variables in the previous regressions.

The linear piecewise (spline) regression applied in this study is:

$$y_{it} = a_i + \beta\overline{thres_{it}} + \gamma' \sum_{j=1}^{J} D_{jit} + \delta' \sum_{j=1}^{J} D_{jit} \times \overline{thres_{it}} + \phi' \mathrm{X}_{it} + country + year \quad (8)$$

, where all the variables are defined as previously, while $\overline{thres_{it}}$ is the cut-point of women participation actual values minus the threshold examined, i.e. 20, 25, 30 per cent and so on. A bandwidth window of ±5 percent is obtained. However, the results for bandwidths of 3 and 4, as well as, for bandwidth of the order of 6-8 and higher remain the same, while lower than 2 per cent there are no observations. After estimating regression (8) a Kernel-weighted linear local polynomial smoothing graph of the predicted-fitted values of (8) on $\overline{thres_{it}}$ is presented. It should be noticed, that higher polynomial order, such as quadratic and cubic terms, present the same concluding remarks. In addition, the structural break tales place in different year for every country as it has been followed in (6) and Tables I and II. In Figures 2-5 the results for thresholds of 20, 25, 29 and 30 per cent respectively are presented. It becomes clear that for thresholds 20 and 25 per cent, there is a “jump” downwards, while there is no difference on the average effect for the threshold of 29 per cent. On the other hand, a jump” upwards is presented for threshold 30 per cent. Therefore, 30 per cent and over result to a significant increase in Public Family Allowances.

Figure 2. Kernel-weighted linear local polynomial smoothing for the 20 per cent. threshold

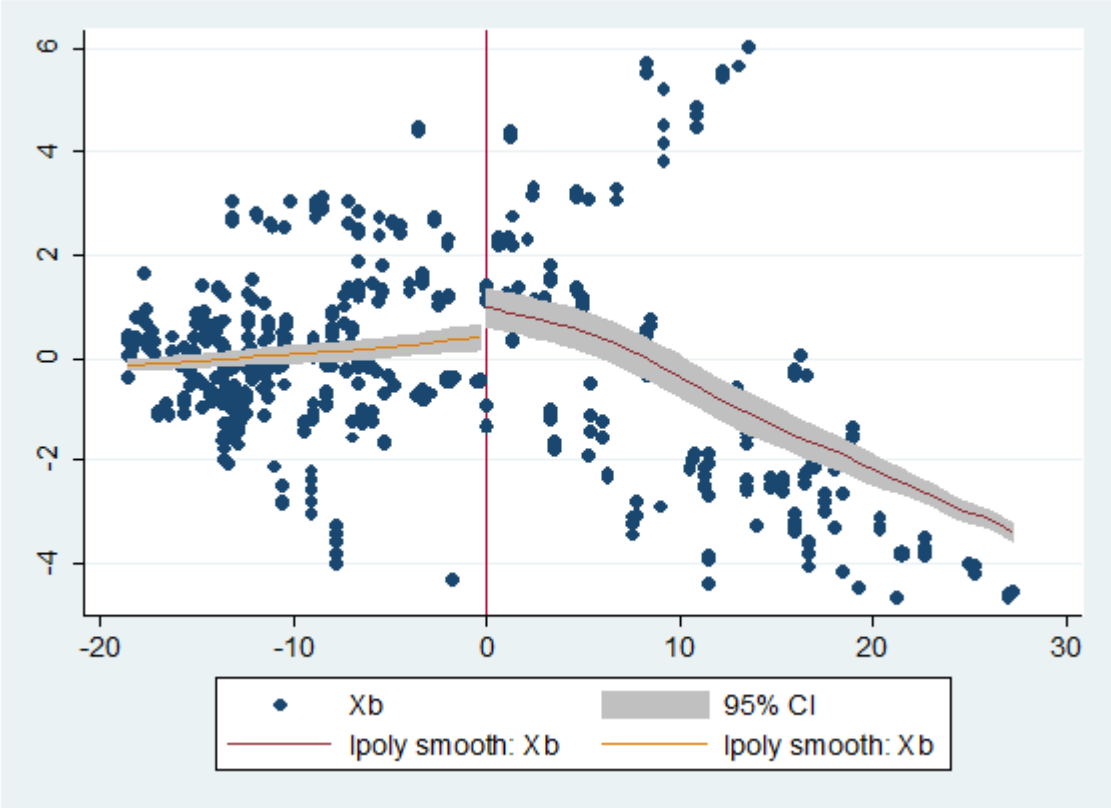


Figure 3. Kernel-weighted linear local polynomial smoothing for the 25 per cent threshold.

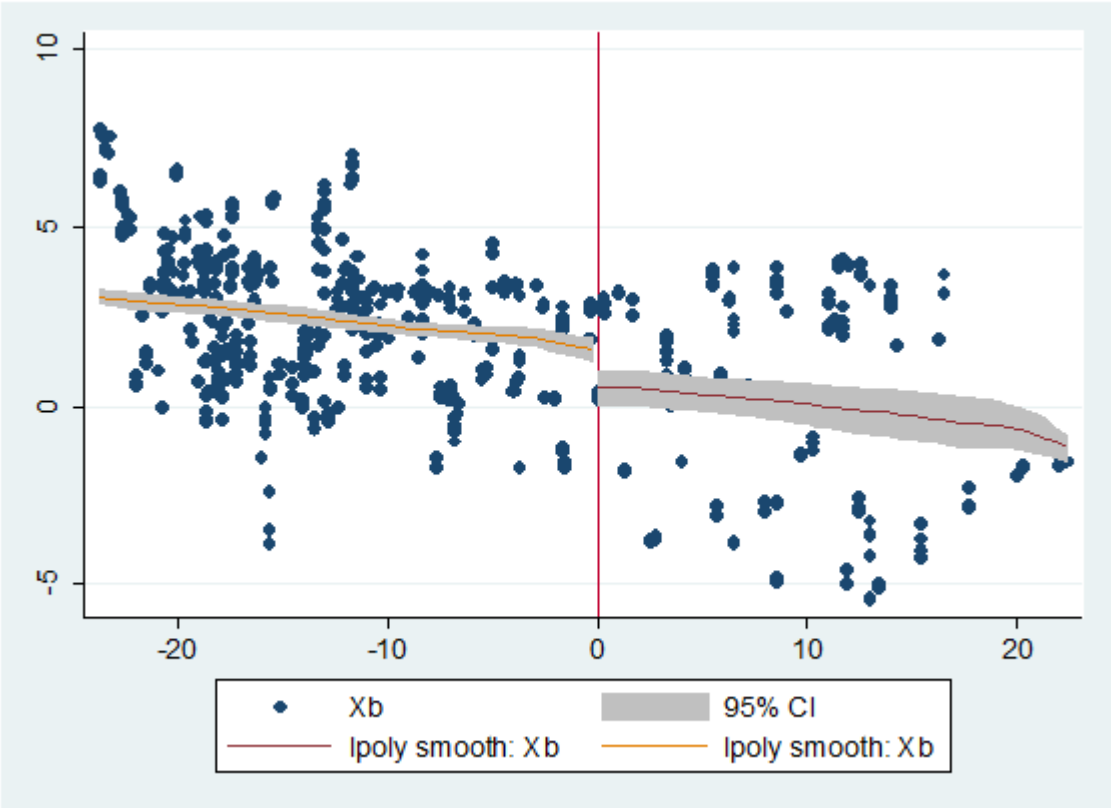


Figure 4. Kernel-weighted linear local polynomial smoothing for the 29 per cent threshold.

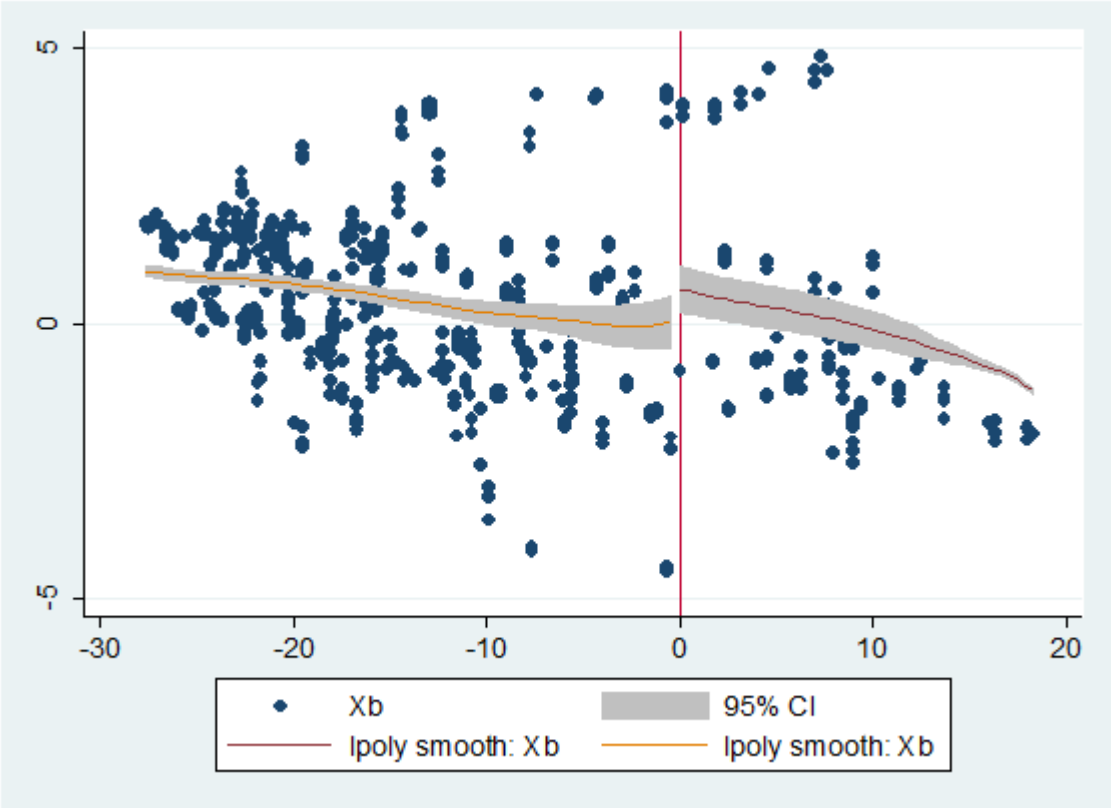
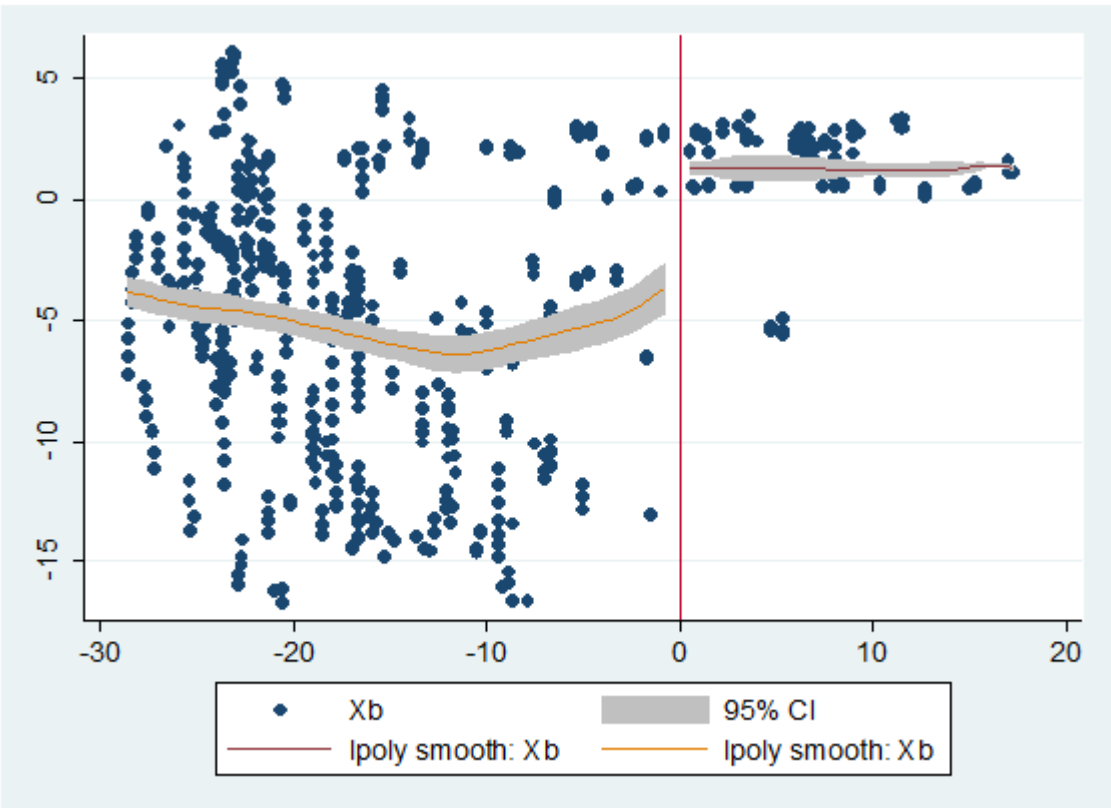


Figure 5. Kernel-weighted linear local polynomial smoothing for the 30 per cent threshold.



2. The two forms of fractionalization, electoral and legislative, for which you control are arguably important factors. However, it is advisable to take into account the institutional structure of the parliamentary systems in exercising oversight over budgetary decisions. Parliamentary systems are typically characterized with strong party discipline and the presence of either a majority or a hung parliament. The literature that you refer to in Section 4: Empirical Robustness also explicitly distinguishes between majority and coalition governments. Given this institutional structure and the literature that you refer to, it is advisable to re-do your empirical analysis by only taking into account the share of female MPs in the caucus of the governing political party for the subset of majority parliaments. For the subset of observations with a hung parliament, you may consider doing your empirical analysis for the caucuses of the governing political parties (if a formal coalition present) and those of all MPs (if not a formal coalition present).

Thank you very much to the referee for the valuable comment that I have a chance to emprically present the fact that female MPs might not be effective in policy-making when electoral incentives are powerful and they take place in majority governments. As referee suggested, I replicated all the regressions for single-majority and coalition governments using the dataset of "*Comparative Political Data Set 1960-2013*. " by Klaus et al., (2015). Replicating all regressions using a database, which is a subset of coalition parliaments, no change occurred in the main results of the paper[2]. However, relevant replications, which are done using a dataset consists of only majority parliaments, show different results. I present the replicated versions of Table 2 and Table 3 in the paper where 'the fraction of female MPs' and the *'threshold 30'* (30 per cent of women over the total parliamentary seats.) are the main independent variables respectively. Table 2A and Table 3A present the replicated results using the subset of coalition governments[3], while Table 2B and Table 3B show the new findings of using majority governments.

[2] To my knowledge, no data unfortunately exists that distinguishes formal and not formal coalition. If there is such a dataset, which I am unaware of, I would be grateful if referee kindly informs me about.

[3] As an example, I present the replicated results for Table 2 and Table 3 of the paper using only the first (base) sample that has the highest number of observations. Replications of the rest of tables using other two samples give similar results.

Table 2A[4]: The Percentage Share of Female Parliamentarians and Public Spending on Family Allowances across OECD Countries (subset of coalition parliaments)

	Without Controls (PANEL A)			With Controls (PANEL D)		
	(1)	(2)	(3)	(1)	(2)	(3)
VARIABLES	OLS	FE	GMM	OLS	FE	GMM
P. of FP	0.0022	0.0137	0.0036	0.0004	0.0075	0.0078*
	(0.0040)	(0.0108)	(0.0033)	(0.0033)	(0.0104)	(0.0044)
Lag(Family Allowances)			0.8870***			0.7071***
			(0.0419)			(0.0942)
Population rate (over 65)				6.3202**	-5.0222	-0.0724
				(2.8115)	(5.6503)	(4.0593)
Population rate (under 15)				7.5814***	5.9103	1.4250
				(2.4661)	(7.6219)	(2.6497)
Total Old-Age Benefits (%GDP)				-0.0742***	0.0363	0.0334**
				(0.0149)	(0.0565)	(0.0164)
Unemployment Rate				0.1080***	0.0259	0.0114
				(0.0140)	(0.0258)	(0.0090)
Log(GDP per capita)				2.0187***	-0.4312	0.2130
				(0.1725)	(0.7867)	(0.3372)
FLFP				0.0106***	0.0083	-0.0034
				(0.0036)	(0.0187)	(0.0082)
Female Education				-0.1183***	0.4685	0.2420
				(0.0317)	(0.5163)	(0.1917)
Observations	245	245	221	245	245	221
R-squared	0.0325	0.7614		0.5434	0.8047	

All standard errors are robust for the arbitrary heteroscedasticity. One, two and three * indicate significance at the 10, 5 and 1% level respectively. FLFP stands for Female Labor Force Participation Rate. P. of FP stands for percentage share of female parliamentarians.

Table 3A: Female Political Representation over the 30% Female Critical Mass Threshold and Public Spending on Family Allowances (subset of coalition parliaments)

PANEL A

	(Without controls)			(With controls)			
	(1)	(2)	(3)	(1)	(2)	(3)	(4)
VARIABLES	OLS	FE	GMM	FE	FE	GMM	GMM
Threshold-30	-0.1112	0.2837***	0.1232***	0.2326***	0.2130***	0.0912***	0.0878***
	(0.0703)	(0.0670)	(0.0234)	(0.0475)	(0.0490)	(0.0157)	(0.0154)
Lag(Family Allowances)			0.7354***			0.6805***	0.6816***
			(0.0685)			(0.0898)	(0.0920)
Pop. R.(above 65)				-5.3464	-3.4766	-1.1716	0.0345
				(5.9098)	(6.0972)	(3.5190)	(4.0639)
Pop.R. (under 15)				4.7422	4.1179	0.8116	0.3055
				(6.1985)	(6.5213)	(2.4710)	(2.5524)
Total Old-Age Benefits (%GDP)				0.0274	0.0313	0.0383**	0.0362**
				(0.0364)	(0.0429)	(0.0154)	(0.0165)
Unemployment R.				0.0236	0.0229	0.0099	0.0081
				(0.0244)	(0.0239)	(0.0097)	(0.0095)
Log(GDP per capita)				-0.5975	-0.6485	-0.1065	-0.0887
				(0.5771)	(0.6568)	(0.3544)	(0.3397)
FLFP					0.0066		-0.0028
					(0.0183)		(0.0086)
Female education					0.3790		0.2347
					(0.4450)		(0.1882)
Observations	245	245	221	245	245	221	221
R-squared	0.0365	0.7744		0.8114	0.8140		

One, two and three * indicate significance at the 10, 5 and 1% level respectively. All standard errors are robust for the arbitrary heteroscedasticity. FLFP stands for Female Labor Force Participation Rate.

[4] Panel Corrected Standard Error Estimates could not be estimated due to the insufficient number of observations.

As seen in Table 3B, once subset of majority parliaments used for the replications, threshold 30 is no longer significant. As referee also indicated, my paper mentions on the relevant literature which supports the fact that larger coalitions tend to be associated with more expenditure, particularly on transfers (see Perotti and Kontopoulos, 1999) since they allow representation of a greater variety of interests of different social groups. Therefore, although they are represented over a certain threshold, it might be expected that female representatives are not effective on policies that women are interested in. Moreover, electoral incentives sometimes would cause candidates to commit to the policy bundle favoured by a majority of voters as previous literature suggested. Therefore, this empirical result also supports the literature that electoral incentives at some degree encourage the elected representative to act according to the interest of the electorate rather than his/her own interest.

Table 2B: The Percentage Share of Female Parliamentarians and Public Spending on Family Allowances across OECD Countries (subset of single majority parliaments)

	Without Controls (PANEL A)			**With Controls (PANEL D)**		
	(1)	(2)	(3)	(1)	(2)	(3)
VARIABLES	OLS	FE	GMM	OLS	FE	GMM
P. of FP	0.0315***	-0.0007	-0.0019	0.0264***	0.0057	0.0071*
	(0.0057)	(0.0079)	(0.0040)	(0.0069)	(0.0041)	(0.0040)
Lag(Family A.)			0.8621***			0.3053**
			(0.0302)			(0.1243)
Pop. R. (over 65)				10.3128***	-12.0426	-14.2561***
				(2.6122)	(7.0070)	(5.1310)
Pop. R. (under 15)				-1.4425	0.3885	-4.1075
				(2.5722)	(5.9373)	(2.5647)
Total Old-Age Benefits (%GDP)				-0.0023	-0.0406	-0.0195
				(0.0429)	(0.0260)	(0.0237)
Unemployment R.				0.0101	0.0062	0.0046
				(0.0112)	(0.0090)	(0.0087)
Log(GDP per capita)				0.6734**	0.0796	0.0516
				(0.3233)	(0.9823)	(0.5206)
FLFP				0.0110	-0.0082	0.0006
				(0.0085)	(0.0140)	(0.0059)
Female education				-0.0460	0.0866	-0.1535
				(0.0453)	(0.2709)	(0.2098)
Observations	180	180	159	180	180	159
R-squared	0.2600	0.8742		0.4053	0.9001	

All standard errors are robust for the arbitrary heteroscedasticity. One, two and three * indicate significance at the 10, 5 and 1% level respectively. FLFP stands for Female Labor Force Participation Rate. P. of FP stands for percentage share of female parliamentarians.

Table 3B. Female Political Representation over the 30% Female Critical Mass Threshold and Public Spending on Family Allowances (subset of single majority parliaments)

PANEL A							
	Without Controls			**With Controls**			
	(1)	(2)	(3)	(1)	(2)	(3)	(4)
VARIABLES	OLS	FE	GMM	FE	FE	GMM	GMM
Threshold-30	0.4048***	0.0021	-0.0028	-0.0548	-0.0187	-0.0291	-0.0243
	(0.0923)	(0.0546)	(0.0357)	(0.0829)	(0.0614)	(0.0587)	(0.0592)
Lag(Family A.)			0.4198***			0.3022**	0.3092**
			(0.1108)			(0.1202)	(0.1284)
Pop. R. (over 65)				-14.2964	-11.1603	-13.1559**	-13.2244**
				(8.2292)	(7.5717)	(5.9199)	(5.4498)
Pop. R. (under 15)				0.8312	1.8841	-1.6914	-1.7635
				(4.2345)	(5.4836)	(1.9090)	(2.0459)
Total Old-Age B.				-0.0304	-0.0375	-0.0219	-0.0207
(%GDP)				(0.0306)	(0.0267)	(0.0234)	(0.0253)
Unemployment R.				0.0060	0.0064	0.0043	0.0046
				(0.0093)	(0.0085)	(0.0095)	(0.0093)
Log(GDP per				0.0886	0.1690	0.0769	0.0814
capita)				(0.9751)	(0.9921)	(0.5275)	(0.5348)
FLFP					-0.0098		-0.0009
					(0.0135)		(0.0053)
Female Education					0.1376		-0.0956
					(0.2620)		(0.2164)
Observations	180	180	159	180	180	159	159
R-squared	0.1112	0.8742		0.8977	0.8993		

One two three * indicate significance at the 10, 5, 1 per cent level respectively. All standard errors are robust to arbitrary heteroscedasticiy. FLFP stands for Female Labor Force Participation Rate.

Finally, as it is expected, If government types (majority, coalition etc.) are controlled, results of the paper remain unchanged as seen at following tables.

Table 2C: The Percentage Share of Female Parliamentarians and Public Spending on Family Allowances across OECD Countries (controlling government type)

	Without Controls (PANEL A)			**With Controls (PANEL D)**		
	(1)	(2)	(3)	(1)	(2)	(3)
VARIABLES	OLS	FE	GMM	OLS	FE	GMM
P. of FP	0.0091***	0.0126*	0.0023	0.0092***	0.0094	0.0056*
	(0.0021)	(0.0071)	(0.0018)	(0.0024)	(0.0075)	(0.0031)
Government Type	-0.0028	-0.0095	-0.0040	-0.0138	-0.0118	-0.0054
	(0.0175)	(0.0122)	(0.0034)	(0.0159)	(0.0113)	(0.0057)
Lag(Family			0.9085***			0.6759***
Allowances)			(0.0180)			(0.0491)
Pop. R. (over 65)				7.9321***	-5.6587	-3.3926*
				(1.5624)	(3.3143)	(1.9746)
Pop. R.(under 15)				9.6228***	0.3496	-0.7512
				(1.0456)	(2.9765)	(1.2866)
Total Old-Age				0.0031	0.0332	0.0347**
Benefits (%GDP)				(0.0100)	(0.0279)	(0.0140)
Unemployment Rate				0.0130	0.0041	-0.0034
				(0.0087)	(0.0124)	(0.0069)
Log(GDP per capita)				1.3531***	-0.6622	-0.0952
				(0.1010)	(0.4222)	(0.1941)
FLFP				-0.0029	0.0098	0.0028
				(0.0026)	(0.0106)	(0.0045)
Female Education				-0.0928***	0.2756	0.1323
				(0.0185)	(0.2748)	(0.1493)
Observations	551	551	494	551	551	494
R-squared	0.0360	0.7674		0.2655	0.7911	

All standard errors are robust for the arbitrary heteroscedasticity. One, two and three * indicate significance at the 10, 5 and 1% level respectively. FLFP stands for Female Labor Force Participation Rate. P. of FP stands for percentage share of female parliamentarians.

Table 3C: Female Political Representation over the 30% Female Critical Mass Threshold and Public Spending on Family Allowances

PANEL A							
	(Without controls)			**(With controls)**			
	(1)	(2)	(3)	(1)	(2)	(3)	(4)
VARIABLES	OLS	FE	GMM	FE	FE	GMM	GMM
Threshold-30	0.0664	0.2370***	0.1002**	0.2394***	0.2253***	0.0977***	0.0947***
	(0.0501)	(0.0679)	(0.0403)	(0.0639)	(0.0652)	(0.0353)	(0.0357)
Lag(Family Allow)			0.6897***			0.6556***	0.6526***
			(0.0485)			(0.0448)	(0.0404)
Pop. R. (over 65)				-3.0622	-3.4498	-2.7642	-2.6388
				(3.4901)	(3.1497)	(1.9090)	(1.7333)
Pop.R. (under 15)				0.6605	0.5429	-0.6911	-0.6796
				(2.3812)	(2.5098)	(1.1412)	(1.1294)
Total OldAge Ben. (%GDP)				0.0338	0.0312	0.0350***	0.0341**
				(0.0257)	(0.0273)	(0.0120)	(0.0133)
Unemployment R.				0.0088	0.0076	-0.0025	-0.0031
				(0.0124)	(0.0117)	(0.0066)	(0.0062)
Log(GDP per capita)				-0.5139	-0.5503	-0.1135	-0.1171
				(0.3566)	(0.3762)	(0.1916)	(0.1748)
Government Type	0.0170	-0.0137	-0.0081	-0.0139	-0.0155	-0.0084	-0.0088
	(0.0168)	(0.0122)	(0.0066)	(0.0114)	(0.0113)	(0.0070)	(0.0071)
FLFP					0.0030		0.0010
					(0.0093)		(0.0043)
Female Education					0.2267		0.1455
					(0.2489)		(0.1459)
Observations	551	551	494	551	551	494	494
R-squared	0.0158	0.7794		0.8015	0.8031		

One, two and three * indicate significance at the 10, 5 and 1% level respectively. All standard errors are robust for the arbitrary heteroscedasticity. FLFP stands for Female Labor Force Participation Rate.

Moreover, all the results presented from Table 2A to Table 3C are robust to second and third samples used in the paper as well.

3.The ideology of a governing political party over the expenditure on family allowance and the share of its nominated (and elected) female candidates might be correlated. It is advisable to isolate for this third factor by possibly including a control variable for the ideological affiliation of the MPs on the right-left political spectrum.

Previous literature indicates that the ideology of the governing party can determine the size of public expenditures and it is argued that left-wing governments expand social expenditures considering child-related issues and social welfare more than right-wing governments. However, due to the nature of panel dataset used at cross-country level for the entire analysis, an individual level variable as the ideological affiliation of the MPs is impossible to integrate into dataset. However, composition of the government on the right left political spectrum could serve for the same aim. Accordingly, Table 3D control for the variable *cabinet composition* which takes value 1 when there is hegemony of right wing parties (and center) and takes value 2 when there is dominance

of right wing (and center) parties. Moreover it takes values 3, 4, 5 when there is balance of power between left and right, dominance of social democratic together with other left parties and hegemony of social democratic together with other left parties respectively. However, controlling for cabinet composition does not create any change in the main results.

Table 3D: Female Political Representation over the 30% Female Critical Mass Threshold and Public Spending on Family Allowances (controlling cabinet composition)

PANEL A

	(Without controls)			(With controls)			
	(1)	(2)	(3)	(1)	(2)	(3)	(4)
VARIABLES	OLS	OLS	GMM	FE	FE	GMM	GMM
Threshold-30	0.0918**	0.0918**	0.0948**	0.2376***	0.2250***	0.0928***	0.0894**
	(0.0436)	(0.0436)	(0.0386)	(0.0697)	(0.0725)	(0.0346)	(0.0352)
Lag(Family Allowances)			0.6924***			0.6537***	0.6510***
			(0.0509)			(0.0489)	(0.0454)
Population rate (over 65)				-3.7199	-4.1041	-2.6564	-2.5718
				(3.7097)	(3.2576)	(2.0051)	(1.8633)
Population rate (under 15)				0.7563	0.6174	-0.5889	-0.5961
				(2.3650)	(2.4737)	(1.1580)	(1.1669)
Total Old-Age Benefits (%GDP)				0.0317	0.0302	0.0353***	0.0347**
				(0.0236)	(0.0265)	(0.0119)	(0.0135)
Unemployment Rate				0.0097	0.0088	-0.0022	-0.0027
				(0.0120)	(0.0114)	(0.0065)	(0.0062)
Log(GDP per capita)				-0.5542	-0.5909	-0.1364	-0.1439
				(0.3787)	(0.3974)	(0.2108)	(0.1956)
Cabinet composition	-0.0017	-0.0017	-0.0036	-0.0177	-0.0174	-0.0049	-0.0049
	(0.0157)	(0.0157)	(0.0041)	(0.0110)	(0.0111)	(0.0040)	(0.0039)
FLFP					0.0031		0.0013
					(0.0093)		(0.0042)
Female Education					0.1692		0.1270
					(0.2487)		(0.1458)
Observations	551	551	494	551	551	494	494
R-squared	0.0144	0.0144		0.8053	0.8063		

One, two and three * indicate significance at the 10, 5 and 1% level respectively. All standard errors are robust for the arbitrary heteroscedasticity. FLFP stands for Female Labor Force Participation Rate.

Moreover, as referee also emphasized that the ideology of a governing political party over the expenditure on family allowance and the share of its female candidates might be correlated. The rise in women's participation in politics that might make women more likely to favour the left and the effect of female politicians on family allowances may depend on the presence of a left-wing government. Therefore I control percentage of posts occupied by left-wing parties to represent the intensity of left-wing government. Table 3E presents the relevant results.

Table 3E: Female Political Representation over the 30% Female Critical Mass Threshold and Public Spending on Family Allowances (controlling the seat percentage of left-wing parties in parliaments.

PANEL A							
	(Without controls)			**(With controls)**			
	(1)	(2)	(3)	(1)	(2)	(3)	(4)
VARIABLES	OLS	FE	GMM	FE	FE	GMM	GMM
Threshold-30	0.0912**	0.2340***	0.0942**	0.2368***	0.2241***	0.0922***	0.0887**
	(0.0438)	(0.0714)	(0.0380)	(0.0693)	(0.0721)	(0.0340)	(0.0345)
Lag(Family Allowances)			0.6940***			0.6565***	0.6539***
			(0.0499)			(0.0477)	(0.0442)
Population rate (over 65)				-3.5033	-3.8935	-2.5413	-2.4543
				(3.6738)	(3.2379)	(2.0005)	(1.8595)
Population rate (under 15)				0.7163	0.5748	-0.5928	-0.6001
				(2.3742)	(2.4859)	(1.1592)	(1.1652)
Total Old-Age Benefits (%GDP)				0.0314	0.0301	0.0352***	0.0346**
				(0.0240)	(0.0268)	(0.0120)	(0.0136)
Unemployment R.				0.0094	0.0085	-0.0023	-0.0028
				(0.0121)	(0.0115)	(0.0065)	(0.0062)
Log(GDP per capita)				-0.5611	-0.5979	-0.1337	-0.1409
				(0.3798)	(0.3973)	(0.2102)	(0.1942)
Left-wing Gov.	-0.0000	-0.0004	-0.0001	-0.0006	-0.0006	-0.0001	-0.0001
	(0.0006)	(0.0005)	(0.0002)	(0.0004)	(0.0004)	(0.0002)	(0.0002)
FLFP					0.0032		0.0013
					(0.0094)		(0.0043)
Female Education					0.1688		0.1263
					(0.2527)		(0.1457)
Observations	551	551	494	551	551	494	494
R-squared	0.0144	0.7796		0.8034	0.8044		

One, two and three * indicate significance at the 10, 5 and 1% level respectively. All standard errors are robust for the arbitrary heteroscedasticity. FLFP stands for Female Labor Force Participation Rate.

The results show that a left-wing government per se does not effect the size of public spending on family allowances and previous results of the paper on the 30 per cent threshold remain unchanged.

4. It is advisable to consider re-writing your introduction by clearly stating the question you are investigating in the opening paragraph and stating your main contribution(s) in the succeeding paragraphs relative to the existing literature. It is advisable to subordinate the current stand of the literature relative to your contribution as the exposition in its current form obscures the contribution(s) of your paper.

Many thanks for the advise about the order of relevant paragraphs that reveal better the contributions of the paper. I will definitely apply them in the final version of the paper.

5. Section 3, which includes substantive and thorough discussion of the merits of and challenges posed by the empirical strategy, is inadequately organized. It is advisable to:

a) split Section 3 into three separate sections: data and descriptive statistics, baseline empirical strategy as well as baseline results and robustness checks;

First referee also had made some useful comments about the organization of the Section 3. I will follow both comments about this Section and reorganize it adequately.

b) relegate the discussion of the merits and drawbacks of estimators other than those of your baseline strategy to the robustness checks section;

Since baseline strategy in Section 3 takes place before the robustness checks in Section 4 and estimations in both applies same estimators, I already discussed the merits and drawbacks of estimators in baseline strategy but not in the robustness checks section. The section 4 for robustness checks do not make any comparison among estimators and do not discuss their merits and drawbacks. It argues only the results of robustness checks, which is necessary in my opinion, for the application of each estimator.

c) expand the section on data and descriptive statistics and tie up the discussion in this section to the question you are investigating as well as to your empirical findings.

Considering the referee's advice, I can expand the section of data and descriptive statistics with a correlation matrix or scatter plots for main variables of interest and tie up the discussion on them with the main question of the paper and the final empirical findings.

6. It is advisable to summarize and include the rationale(s) for including the listed control variables on page 9 that are commonly used in the literature.

Referee is right that I rationalized the use of control variables quite short and did not explain them in details since previous literature already mentioned about them in their studies. However, I will give more information about them in the final version of the paper.

7. It is not clear why "... the persistent under-representation of women in the OECD parliaments might still be an obstacle for their efficiency in policy decision making on public family allowances..." This final sentence of the paper introduces a normative perspective on the issue that has not addressed in the analysis of the paper.

As the OECD average of the percentage share of female parliamentarians have been still under 30 per cent since many years, I argued with this last sentence that this underrepresentation might still be an obstacle for their efficiency on public family allowances. It was an underpinning phrase to support the main finding of the paper about 30 per cent threshold. However, if it is found so normative, I agree that it should be removed as well.

REFERENCES

Armingeon, Klaus, Christian Isler, Laura Knöpfel and David Weisstanner. 2015. *Supplement to the Comparative Political Data Set – Government Composition 1960-2013*. Bern: Institute of Political Science, University of Berne.

Andrews, D. W. K. (1993). Tests for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica*, 61(4), 821-856.

Caughey, D.M. and Sekhon, J.S. (2011). Elections and the regression discontinuity design: lessons from close US house races, 1942-2008. Polit. Anal. 19, 385–408.

Imbens, G.W. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2), 615-635.

Kontopoulos, Y., and Perotti, R. (2008). Government Fragmentation and Fiscal Policy Outcomes: Evidence from OECD Countries. *Fiscal Institutions and Fiscal Performance*, 81.

Lee, D. and Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, 48(2), 281-355.

Thistlethwaite, D. L. and Campbell D. T. (1960). Regression-Discontinuity Analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51(6), 309-317.