The effect of the ratio of salaried female employees on maternal mortality

1. QUESTION

Being the 5th Millennium Development Goal, the relationship of maternal mortality with various macrostructural factors has been extensively studied in the last decades. Sajedinejad et al. (2015) showed that decreasing maternal mortality requires dealing with factors such as political will, reallocation of national resources (especially health resources) in the governmental sector, education, attention to the expansion of the private sector trade and improving spectrums of governance. Even though 28-40% of maternal deaths could be preventable (Clark et al., 2008), maternal mortality reduction programs built around the above mentioned indicators have not been completely successful (Sajedinejad et al., 2015).

Several studies found composite indexes such as HDI (Human Development Indicator) and GDI (Gender Development Indicator) to be the most powerful predictors of variation in maternal mortality rates, together with infant mortality rates and total fertility rates, which were also found to be highly predictive of maternal mortality variation across countries (McAlister and Baskett, 2006). Research has also consistently observed an inverse correlation between women's education level and maternal mortality in the developing world (Koch et al., 2012).

In spite of describing a large part of the differences in the general quality of life between countries, in previous research these macro-level indicators and composite indexes on healthcare, education and gender equality still could not explain all the variation of maternal mortality. This made me think about other, more individually driven aspects of gender equality. I identified labor market participation as a potential factor that effects maternal mortality, which resulted in the following research question:

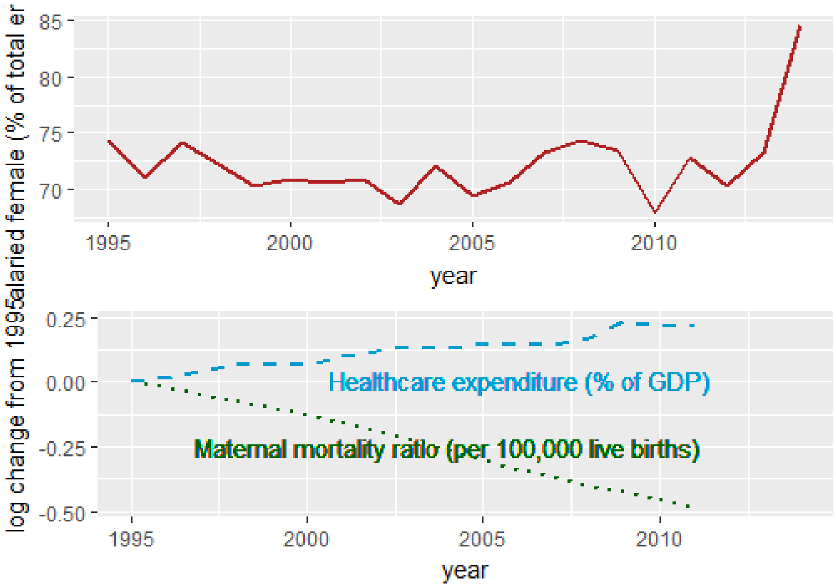
What is the effect of the ratio of salaried female employees on maternal mortality?

I expect this question to provide input on a more holistic dilemma: How much difference does it make in the health prospects of woman to be employed as a wage or salaried worker? Compared to the main alternatives of being a family worker (stay-at-home mothers) or to be employed in the family business (which in many cases means working in the agricultural sector), salaried workers are relatively more independent socially and financially.

2. DATA

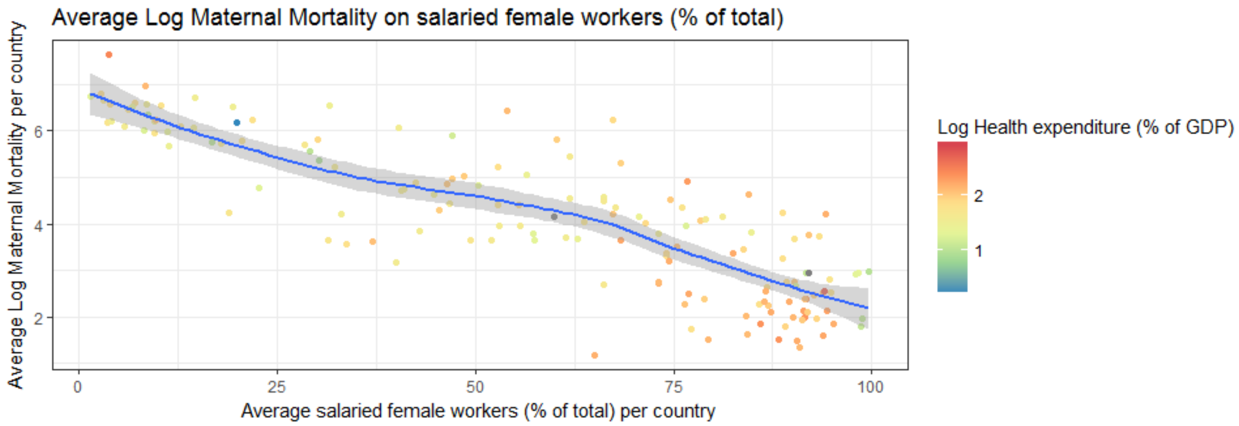
The data was collected from the World Bank's World Development Indicators website through the WDI API. After merging the original dataset it had 4340 observations and 7 variables. The original variables were Country name, Year, Wage and salaried workers, female (% of females employed), Maternal mortality ratio (per 100,000 live births), Health expenditure, total (% of GDP), GDP per capita, PPP (constant 2005 international $) and Population, total. The time period for the dataset was from 1995 to 2014. Population, total was converted into million people and only observations with years between 1995 and 2014 were kept. The dataset at this point was unbalanced, which required the further analysis of the distribution of missing observations.

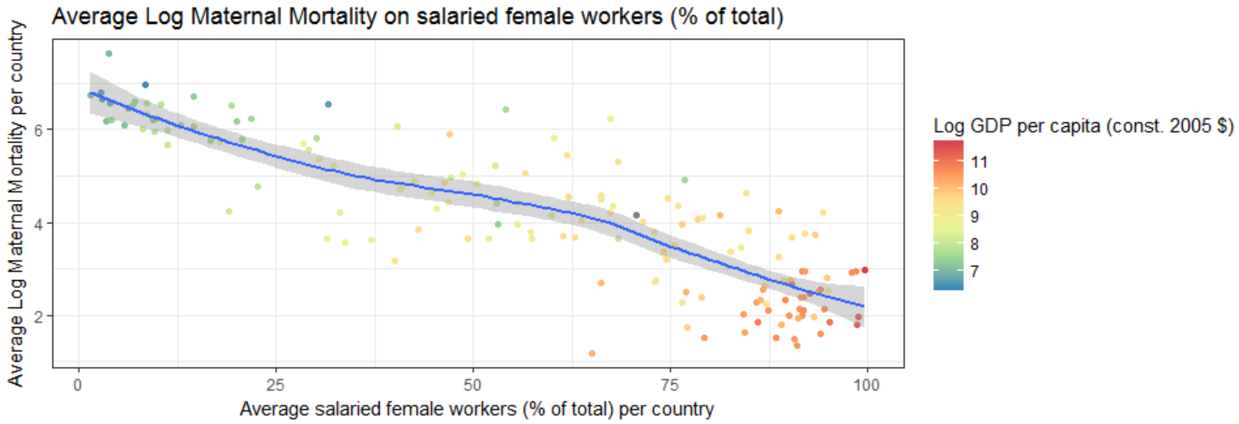
Yearly global averages of Maternal mortality ratio, of the ratio of Salaried workers of all females employed and of Healthcare expenditure (% of GDP) were plotted. It is visible on the global trend of all countries that maternal mortality ratio decreased by 41% (from 255 to 150 per 100,000 live births), ratio of Salaried workers of all females employed increased by 13%, and Healthcare expenditure (% of GDP) increased by almost 25% in the period of 1995 to 2014.



Further description of the dataset is to be found in Appendix I, which also includes the descriptive statistics of the used variables and comparable trends of selected countries of interest.

The average relationship between Log maternal mortality and the Ratio of salaried female workers was visualized with a loess regression. On the first plot, scatterplots were colored based on Log health expenditure values, while on the second plot, Log GDP per capita was used for coloring.





It is visible on both scatterplots that in the 1995-2014 period average Log Maternal mortality was higher in countries where average Ratio of salaried female workers was higher, on average. This does not necessarily mean though that an increase in the Ratio of salaried female workers led to higher Log Maternal mortality on the short or longer term. There might have been other events and other differences in countries that had an effect on Maternal mortality. The coloring provides evidence that Log GDP per capita is a potential confounder of the average relationship between Log Maternal mortality and Ratio of salaried female workers, while Log Health expenditure does not show the same trend, as countries with higher and lower Log Healthcare expenditure values are distributed all across higher and lower values of Maternal mortality and ratio of salaried female workers.

3. DISCUSSION

Throughout the entire analysis standard error estimates were adjusted to ensure robustness of all regression estimates. In the regressions countries were not weighted based on their population, since both Maternal mortality ratio (per 100,000 live births), and Ratio of salaried workers of females employed were defined by default as ratios, and GDP per capita was normalized by population.

To identify the regression that captures the effect I am after, I estimated multiple OLS, fixed effect and first different regressions. Multiple OLS regressions were estimated for three different years (1995, 2007 and 2014), with and without GDP per capita and health expenditure as control variables. The results on the OLS regression estimates can be found in Appendix II.

The fixed effect model also estimated with explicit time dummies. Again, standard error estimates were adjusted to ensure robustness of all regression estimates. The four estimated fixed effect models provided similar results in magnitude, all coefficients being statistically significant at the 5% and most on the 1% level even. Results of the FE regressions can be found in Appendix III.

In FE1, comparing two countries that have different ratio of Salaried female workers relative to its mean in country i, but are the same in everything else that does not change in time, Maternal mortality is expected to be higher by 4.00%, on average, relative to its mean value in country i, where or when Ratio of salaried female workers is higher by 1 percentage point than its long-term average in counry i. In FE4, comparing two countries that have different ratio of Salaried female workers relative to its mean in country i, have the same level of Log GDP per capita, Log Healthcare expenditure and Log Population relative to their mean in country i, but are the same in everything else that does not change in time, Maternal mortality is expected to be higher by 1.00%, on average, relative to its mean value in country i, where or when Ratio of salaried female workers is higher by 1 percentage point than its long-term average in country i.

For the first differences model, Log Maternal mortality and Ratio of salaried female workers were added to the regression both without lags and with 1 to 6 lags, where regressions were all estimated with explicit time dummies. Results of the FD regressions can be found in Appendix IV. The largest cumulative coefficient belonged to the FD estimate with 6 lags, which had a cumulative beta of 1.26, however none of the coefficients were significant statistically and the regression has an R-squared of 1. The cumulative coefficient of the rest of the regressions is almost the same: -0.04 for FD1 and FD3 and -0.045 for FD2 with R-squared values between 0.36 and 0.48. The coefficient of the contemporaneous effect is statistically significant at the 1% level for FD1 and FD2, being -0.04 and -0.03 respectively. More of the year dummies are statistically significant at the 5% level for FD2 than for FD1, and only year dummy 2000 is statistically significant at the 1% level, in the FD2 regression. This is probably due to an extreme event in that year which influenced Maternal mortality significantly in our panel.

In FD1 with no lags the coefficient of the contemporaneous right-hand-side variable is -0.04. In years when Ratio of salaried female workers increases by 1 percentage point more, Maternal mortality decreases by 4.00% more, on average. The cumulative coefficient is -0.04, which means that comparing countries, or years, with a 1 percentage point difference in Ratio of salaried female workers, Maternal mortality decreases by 4.00% more on average in the year in where, or when, Ratio of salaried female workers increases by 1 percentage point more.

In FD2 with two lags the coefficient of the contemporaneous right-hand-side variable is -0.03. In years when ratio of salaried female workers increases by 1 percentage point more, Maternal mortality decreases by 3.00% more, on average. The average decrease in Maternal mortality is 1.00% the year after. The average decrease two years later is 0.5%. The cumulative coefficient is -0.045, which means that comparing countries, or years, with a 1 percentage point difference in Ratio of salaried female workers, Maternal mortality decreases by 4.50% more on average after two years in where, or when, Ratio of salaried female workers increases by 1 percentage point more.

Adding control variables to the FD regressions we find that for FD with no lag the coefficient of the contemporaneous right-hand-side variable becomes statistically not significant, while for FD2, the coefficient changes from -0.03 without controls to -0.05 with controls, being statistically significant at the 1% level. The cumulative coefficient is -0.01 for FD1 with controls, and it is -0.104 for FD2 with controls. In FD2 the coefficient of the first lag is also statistically significant at the 1% level. These results suggest that Log GDP per capita and Log Health expenditure are confounders in the relationship, and that based on the FD regression most of the effect of the change in the Ratio of wage and salaried female employees on Maternal mortality happens either in the contemporaneous year or the next year, depending on whether we add controls to the regression. Results of the FD regressions with controls can be found in Appendix V.

Comparing multiple models

	Dependent variable: Log Maternal mortality			
	OLS2007 with controls	FE4 with controls	FD2 no controls	FD2 with controls
	(1)	(2)	(3)	(4)
salaried	-0.01	-0.01**		
	(0.01)	(0.005)		
diff(salaried)			-0.03***	-0.05***
			(0.01)	(0.01)
stats::lag(diff(salaried), 1:2)1			-0.01	-0.05***
			(0.02)	(0.02)
stats::lag(diff(salaried), 1:2)2			-0.005	-0.004
			(0.01)	(0.01)
Constant	14.42***		-0.94**	0.40**
	(1.22)		(0.38)	(0.19)
Observations	90	1,677	164	111
R^2	0.77	0.74	0.36	0.80
Note:	*p<0.1; **p<0.05; ***p<0.01			

Estimates cannot be interpreted as the measure of the causal effect unless it is assumed that reverse causality is not an issue and the fixed effects take care of all kinds of selection and other confounders. Cross-sectional FE do take care of selection on fixed country characteristics (countries with more income select themselves to run initiatives to lower Maternal mortality). Time FE takes care of trends in both Ratio of salaried workers and Maternal mortality as long as these trends are the same across countries, but might miss potential time varying selection and country-specific trends.

When including control variables FD and FE coefficient estimates are not very different in mangitude, however, the FE coefficient is closer to the OLS. The best estimate is the one for which the common trends assumption is the most likely to hold. FD looks at changes happening in certain years (same year or a subsequent year), but might miss the effect of the change in other years. FE looks at levels not changes relative to the long-term mean in a country. This has the advantage that fixed cross-country differences don't confound the estimates, and that also long-term effects are captured. Looking at levels also has the disadvantage that it needs stationary series for the time series regression run in each country.

Based on these arguments I identified FE as the best estimate for the current question, because it is the one for which the common trends assumption is the most likely to hold (fixed cross-country differences don't confound the estimates, and that also long-term effects are captured).

To uncover unobserved heterogenity I created a low- and high-income group between countries based on the median of GDP per capita, which is 8,514 USD. Results of the regressions on group datasets and analysis of results can be found in Appendix VI. We can conclude that some unobserved heterogenity is still present in the previous FD and FE estimates of the full panel when controlling for the above mentioned variables, however, the low- and high-income groups faced a very similar effect in magnitude, therefore the previous estimates can be considered to be robust.

4. RESULTS

Overall, we can say that Ratio of salaried female workers has an effect on Maternal mortality, which is statistically significant also when controlling for Log GDP per capita and Log Health expenditure. The slope coefficients of the contemporaneous effect in the unbalanced panel were statistically significant at the 5.00% level both in the fixed effects and in the first difference model. When adding controls, the slope coefficient of the first lag was statistically significant at the 1.00% level in the first difference regression.

These results mean that most of the effect between these variables takes place immediately or in the first year following the change in the explanatory variable. Even by adding control variables commonly used in former research on the topic, we could not explain the relationship fully in this analysis, but as a conclusion we can still say that very likely there is a causal relationship between the two variables.

To understand the causal mechanism fully, we would need to include all potential explanatory variables on gender equality, healthcare, economic growth, education, labor market, etc. By examining unobserved heterogeneity in the low-income and high-income groups of countries we could conclude that there is a slight difference between the two groups, however, the coefficient estimates on the full panel can still be considered robust. For future policy making, based on the results of this study, it would be reasonable to examine country-specific trends and policy examples to understand the effect of Ratio of Salaried female workers (% of employed) on Maternal mortality in details. Based on the panel between 1995 and 2014 we can confirm that the ratio of females employed as salaried workers has a visible effect on Maternal mortality ratios, therefore it has an added value in explaining the effect to look into more specific indicators beyond the holistic composite indexes usually used.

5. REFERENCES

Sajedinejad, S., Majdzadeh, R., Vedadhir, A., Tabatabaei, MG., Mohammad, K. (2015) Maternal mortality: a cross-sectional study in global health. Global Health. 2015;11:4.

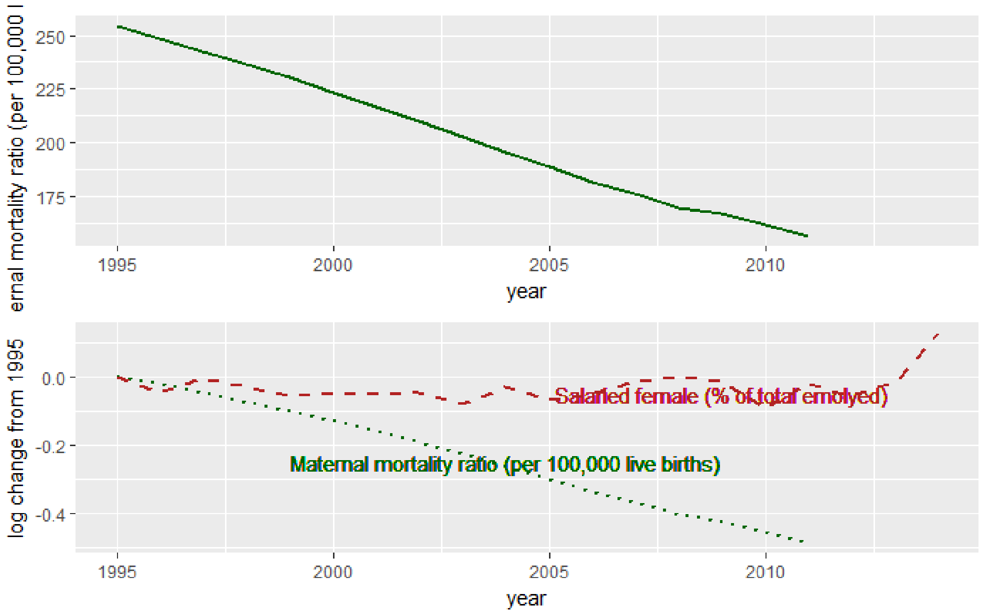
Clark, SL., Belfort, MA., Dildy, GA., Herbst, MA., Meyers, JA., Hankins, GD. (2008) Maternal death in the 21st century: causes, prevention, and relationship to cesarean delivery. Am J Obstet Gynecol. 2008 Jul; 199(1):36.e1-5; discussion 91-2. e7-11.

McAlister, C., Baskett, T. F. (2006) Female Education and Maternal Mortality: A Worldwide Survey. Am J Obstet Gynaecol. 2006;28(11):983–990

Koch, E., Thorp, J., Bravo, M., Gatica, S., Romero, CX., et al. (2012) Women's Education Level, Maternal Health Facilities, Abortion Legislation and Maternal Deaths: A Natural Experiment in Chile from 1957 to 2007. PLOS ONE 7(5): e36613. doi: 10.1371/journal.pone.0036613

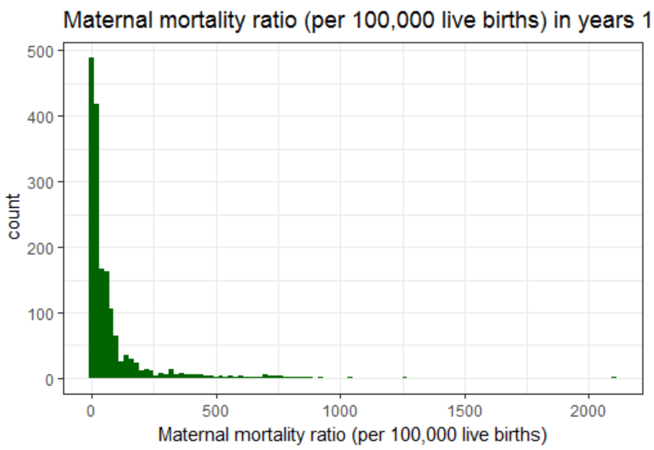
APPENDIX *I.* | Description of the dataset

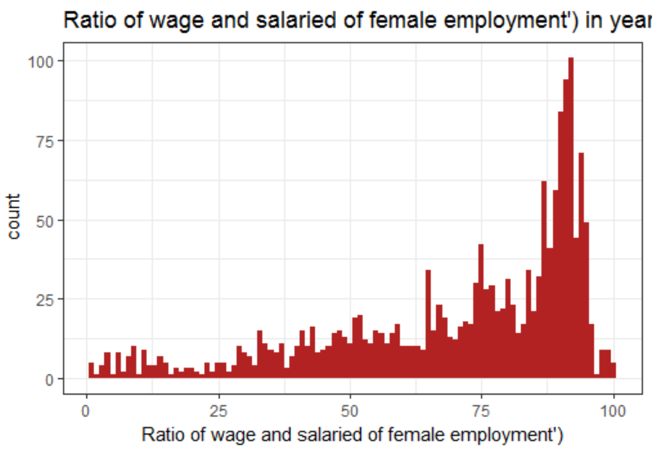
Yearly global averages of Maternal mortality ratio, of the ratio of Salaried workers of all females employed and of Healthcare expenditure (% of GDP) were plotted. It is visible on the global trend of all countries that maternal mortality ratio decreased by 41% (from 255 to 150 per 100,000 live births) and ratio of Salaried workers of all females employed increased by 13% in the period of 1995 to 2014.

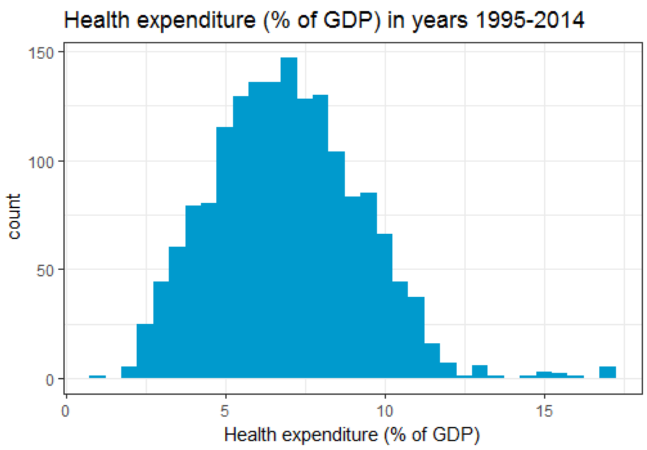


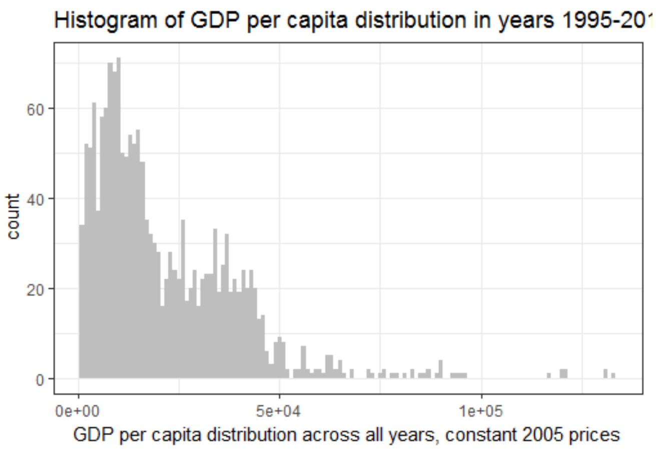
Distributions of observations with missing values for the Log Maternal mortality, Ratio of salaried female workers, Log GDP per capita and Log Health expenditure variables were visualized in histograms.

For Log Maternal mortality more than 30 observations had 20 missing values, for Ratio of salaried female workers the number of observations with missing values was the highest for more than 18 values, similarly to Log GDP per capita and Healthcare expenditure.







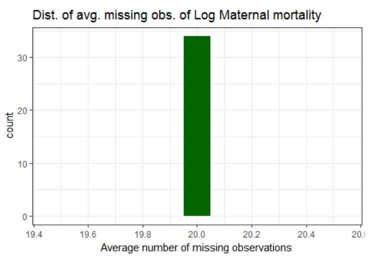


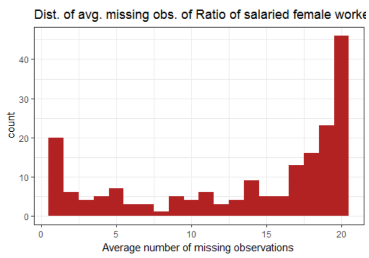
Differences for GDP per capita, Maternal mortality ratio and Healthcare expenditure make more sense in relative than in absolute terms in the current context of the main question. Therefore log variables were created for these variables.

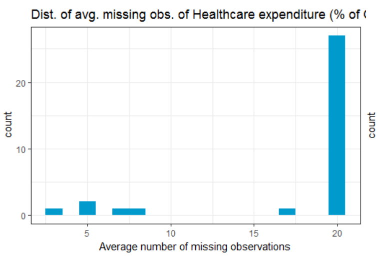
Descriptive Statistics: transformed and original variables in the unbalanced panel

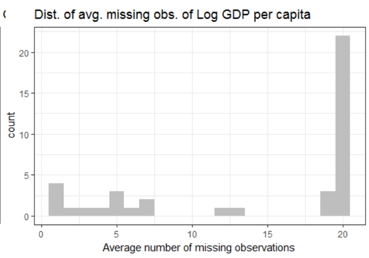
Statistic	N	Mean	St. Dev.	Min	Max
health	3,755	6.259	2.680	0.368	30.829
mmort	3,660	236.631	334.648	3	2,900
pop	4,334	29.788	120.847	0.009	1,364.270
salaried	1,808	71.850	23.641	1.000	99.900
gdppc	3,770	15,656.930	18,989.610	246.671	137,164.400
lngdppc	3,770	8.976	1.243	5.508	11.829
lnmmort	3,660	4.342	1.663	1.099	7.972
lnpop	4,334	1.211	2.417	-4.686	7.218
lnhealth	3,755	1.747	0.425	-0.999	3.428

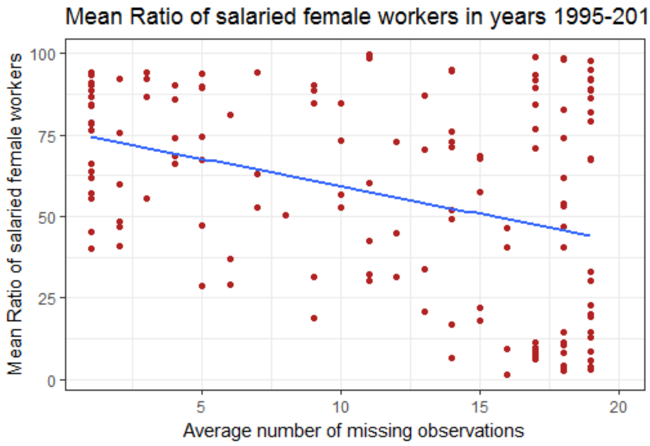
By plotting the number of observations with missing values and the average country data in the given time period we can see that for both Log Maternal mortality and Log Health expenditure the countries with more missing values tend to have higher mean values of the variables, however, the countries with higher Log GDP per capita tend to have more observations with missing values. Overall, we can say that the dataset is moderately unbalanced.

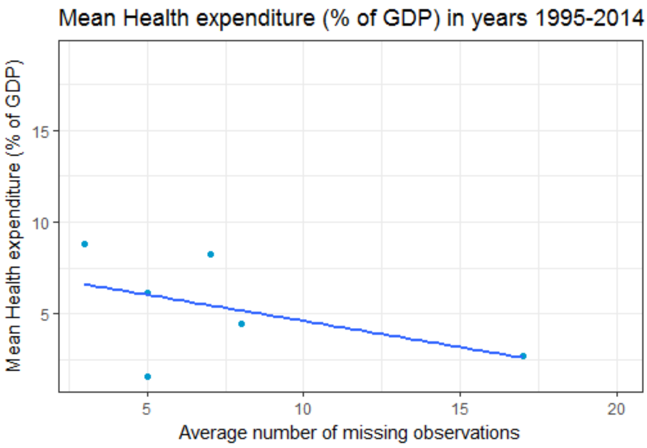


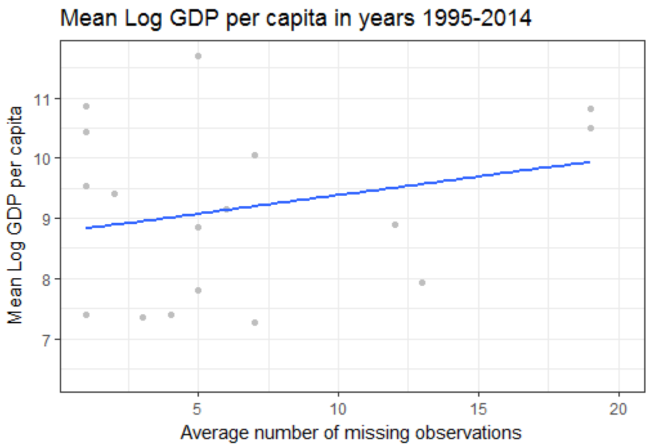







Mean Ratio of salaried female workers in years 1995-2014
Mean Ratio of salaried female workers
100
75
50
25
0
5
10
15
20
Average number of missing observations


Mean Health expenditure (% of GDP) in years 1995-2014
Mean Health expenditure (% of GDP)
15
10
5
5
10
15
20
Average number of missing observations


Mean Log GDP per capita in years 1995-2014
Mean Log GDP per capita
11
10
9
8
7
5
10
15
20
Average number of missing observations

Selecting countries of interest with a low number of missing observations we can see that individual countries experienced very different trends between 1995 and 2014. Comparing El Salvador and Turkey, who had comparable levels of Salaried workers ratio amongst female workers, and similar Maternal mortality in the given period, experienced very different trend of Healthcare expenditure. In Turkey healthcare spending has increased while El Salvador the same decreased, to finally reach a 5-7% treshold for both.

APPENDIX *II.* | OLS regression estimates

In OLS1995, comparing two countries, the country with a 1 percentage point higher Ratio of salaried female workers was expected to have a 4.00% lower Maternal Mortality, on average. In OLS2007, comparing two countries, the country with a 1 percentage point higher Ratio of salaried female workers was expected to have a 4.00% lower Maternal Mortality, on average. In OLS2014, comparing two countries, the country with a 1 percentage point higher Ratio of salaried female workers was expected to have a 5.00% lower Maternal Mortality, on average.

In OLS2007 with controls, comparing two countries with the same Log Health expenditure and Log GDP per capita, the country with a 1 percentage point higher Ratio of salaried female workers was expected to have a 1.00% lower Maternal Mortality, on average. Despite of the slope coefficients of control variables being statistically significant at the 1% level, the slope coefficient on Maternal Mortality was not statistically significant at the 5% level in the year 2007. The OLS regressions provided results of similar magnitude, even though by adding potential confounders, the effect of the Ratio of salaried female workers on Maternal mortality became significantly smaller on average.

OLS models

	Dependent variable: Maternal Mortality			
	lnmmort			
	ols1995	ols2007	ols2014	ols2007c
	(1)	(2)	(3)	(4)
salaried	-0.04^{***}	-0.04^{***}	-0.05^{***}	-0.01
	(0.01)	(0.003)	(0.01)	(0.01)
lnhealth				-0.82^{***}
				(0.22)
lngdppc				-0.92^{***}
				(0.15)
Constant	6.61^{***}	6.33^{***}	6.70^{***}	14.42^{***}
	(0.46)	(0.25)	(1.22)	(1.22)
Observations	66	93	42	90
R^2	0.43	0.61	0.27	0.77
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$			

APPENDIX *III.* | FE regression estimates

Comparing FE models with expicit time dummies

	Dependent variable: Log Maternal Mortality			
	FE1	FE2	FE3	FE4
	(1)	(2)	(3)	(4)
salaried	-0.04***	-0.02***	-0.04***	-0.01**
	(0.003)	(0.005)	(0.003)	(0.005)
lngdppc		-0.85***		-0.88***
		(0.13)		(0.12)
lnhealth			-0.71***	-0.71***
			(0.19)	(0.18)
lnpop				0.01
				(0.04)
Observations	1,725	1,715	1,687	1,677
R^2	0.58	0.71	0.61	0.74
Note:	*p<0.1; **p<0.05; ***p<0.01			

APPENDIX *VI.* | FD regression estimates

Comparing FD models with expicit time dummies

	Dependent variable: Log Maternal Mortality			
	FD1	FD2	FD3	FD4
	(1)	(2)	(3)	(4)
diff(salaried)	-0.04***	-0.03***	-0.01	1.12
	(0.004)	(0.01)	(0.02)	
stats::lag(diff(salaried), 1:2)1		-0.01		
		(0.02)		
stats::lag(diff(salaried), 1:2)2		-0.005		
		(0.01)		
year1996	0.04	0.67		
	(0.17)	(0.82)		
year1997	-0.22	0.67*		
	(0.19)	(0.37)		
stats::lag(diff(salaried), 1:4)1			0.01	
			(0.01)	
stats::lag(diff(salaried), 1:4)2			0.01	
			(0.03)	
stats::lag(diff(salaried), 1:4)3			0.01	
			(0.02)	
stats::lag(diff(salaried), 1:4)4			0.02	
			(0.01)	
year1998	-0.08	0.54	0.11	
	(0.15)	(0.41)	(0.53)	
year1999	-0.10	0.54	0.10	
	(0.15)	(0.37)	(0.57)	
year2000	0.03	1.59***		
	(0.18)	(0.58)		
year2001	-0.25	0.91*	-0.30	
	(0.21)	(0.51)	(0.82)	
year2002	-0.37*	0.69	-0.56	
	(0.21)	(0.50)	(0.62)	
year2003	-0.03	1.06**	0.10	
	(0.18)	(0.51)	(0.50)	
stats::lag(diff(salaried), 1:6)1				1.25

stats::lag(diff(salaried), 1:6)2				-0.16
stats::lag(diff(salaried), 1:6)3				-0.73
stats::lag(diff(salaried), 1:6)4				-0.02
stats::lag(diff(salaried), 1:6)5				0.14
stats::lag(diff(salaried), 1:6)6				-0.34
year2004	-0.20	0.99**	-0.56	3.35
	(0.19)	(0.45)	(0.54)	
year2005	-0.08	0.98*	-0.44	
	(0.18)	(0.53)	(0.42)	
year2006	-0.16	0.88**	-0.57	0.14
	(0.19)	(0.41)	(0.52)	
year2007	-0.34*	0.84**	-0.59	
	(0.19)	(0.40)	(0.52)	
year2008	-0.22	0.91**	-0.63	
	(0.20)	(0.44)	(0.52)	
year2009	-0.18	0.57	-0.06	
	(0.20)	(0.47)	(0.15)	
year2010	0.04	1.20*	1.34	
	(0.21)	(0.63)	(1.58)	
year2011	-0.13	1.22**	-0.98	
	(0.22)	(0.54)	(0.76)	
year2012	-0.37*	0.78*	-0.45	
	(0.22)	(0.46)	(0.65)	
year2013	-0.47**	0.96*		
	(0.23)	(0.53)		
year2014	-0.08			
	(0.29)			
Constant	0.27	-0.94**	0.43	23.56
	(0.21)	(0.38)	(0.56)	
Cumulative Coeff	-0.04	-0.045	-0.04	1.26
Observations	687	164	44	10
R^2	0.48	0.36	0.40	1.00
Note:	*p<0.1; **p<0.05; ***p<0.01			

APPENDIX *V.* | FD regression estimates with controls

Comparing FD models with expicit time dummies and controls

	Dependent variable: Log Maternal Mortality		
	FD5	FD6	FD7
	(1)	(2)	(3)
diff(salaried)	-0.01	-0.05***	0.03
	(0.01)	(0.01)	
stats::lag(diff(salaried), 1:2)1		-0.05***	
		(0.02)	
stats::lag(diff(salaried), 1:2)2		-0.004	
		(0.01)	
stats::lag(diff(lnhealth), 1:2)1		1.04**	
		(0.44)	
stats::lag(diff(lnhealth), 1:2)2		-0.14	
		(0.31)	
stats::lag(diff(lngdppc), 1:2)1		1.07***	
		(0.32)	
stats::lag(diff(lngdppc), 1:2)2		-0.08	
		(0.22)	
stats::lag(diff(lnpop), 1:2)1		-0.04	
		(0.10)	
stats::lag(diff(lnpop), 1:2)2		-0.20*	
		(0.10)	
year1996	-0.12	-0.18	
	(0.14)	(0.19)	
year1997	-0.25	-0.64**	
	(0.16)	(0.30)	
year1998	-0.13	-0.06	
	(0.15)	(0.30)	
stats::lag(diff(salaried), 1:4)1			-0.03
stats::lag(diff(salaried), 1:4)2			-0.01
stats::lag(diff(salaried), 1:4)3			0.01
stats::lag(diff(salaried), 1:4)4			0.02
stats::lag(diff(lnhealth), 1:4)1			0.05
stats::lag(diff(lnhealth), 1:4)2			-1.00

stats::lag(diff(lnhealth), 1:4)3			-0.77
stats::lag(diff(lnhealth), 1:4)4			-0.63
stats::lag(diff(lngdppc), 1:4)1			-0.54
stats::lag(diff(lngdppc), 1:4)2			-3.13
stats::lag(diff(lngdppc), 1:4)3			1.07
stats::lag(diff(lngdppc), 1:4)4			0.10
stats::lag(diff(lnpop), 1:4)1			-3.62
stats::lag(diff(lnpop), 1:4)2			-5.21
stats::lag(diff(lnpop), 1:4)3			-4.55
stats::lag(diff(lnpop), 1:4)4			0.54
year1999	-0.18	-0.09	-0.01
	(0.15)	(0.29)	
year2000	-0.03	-0.25	
	(0.16)	(0.25)	
year2001	-0.19	-0.01	-0.05
	(0.17)	(0.25)	
year2002	-0.33*	-0.24	-0.20
	(0.18)	(0.27)	
year2003	-0.12	0.004	-0.19
	(0.16)	(0.30)	
year2004	-0.23	0.26	
	(0.17)	(0.38)	
year2005	-0.12	-0.45*	
	(0.14)	(0.24)	
year2006	-0.25*	0.04	
	(0.14)	(0.32)	
year2007	-0.34**	0.02	
	(0.15)	(0.27)	
year2008	-0.22	-0.10	
	(0.15)	(0.45)	
year2009	-0.14	-0.72**	
	(0.15)	(0.35)	
year2010	-0.09	0.46	
	(0.15)	(0.42)	
year2011	-0.28**	0.47	

	(0.14)	(0.49)	
year2012	-0.40**	-0.20	
	(0.15)	(0.28)	
year2013	-0.38**	0.19	
	(0.16)	(0.28)	
year2014	0.26		
	(0.23)		
diff(lnhealth)	-0.42**		
	(0.20)		
diff(lngdppc)	-0.97***		
	(0.19)		
diff(pop)	0.001		
	(0.001)		
Constant	0.30**	0.40**	0.58
	(0.13)	(0.19)	
Cumulative Coeff	-0.01	-0.104	0.02
Observations	646	111	22
R^2	0.69	0.80	1.00
Note:	*p<0.1; **p<0.05; ***p<0.01		

APPENDIX *VI.* | Regression estimates on datasets of country groups

To uncover unobserved heterogenity I created a low- and high-income group between countries based on the median of GDP per capita, which is 8,514 USD. By creating group 1 of low-income countries (GDP per capita smaller or equal to 8,500 USD, contant 2005 prices) and group 2 for high-income countries, the difference in the mechanism between Ratio of Salaried female workers and Maternal mortality changes for low- and high-income countries became visible.

Comparing multiple models for high- and low-income countries

	Dependent variable: Log Maternal mortality			
	FD2c, G:1	FE4c, G:1	FD2c, G:2	FE4c, G:2
	(1)	(2)	(3)	(4)
diff(salaried)	-0.03		-0.03	
			(0.02)	
stats::lag(diff(salaried), 1:2)1	-0.02		-0.06^{**}	
			(0.02)	
stats::lag(diff(salaried), 1:2)2	0.01		-0.03^{**}	
			(0.01)	
salaried		-0.02^{***}		-0.002
		(0.005)		(0.01)
Constant	4.26		0.37	
			(0.40)	
Observations	13	423	170	1,254
R^2	1.00	0.59	0.32	0.54
Note:	*p<0.1; **p<0.05; ***p<0.01			

Dataset	Full panel	Full panel	Group 1	Group 1	Group 2	Group 2
Regression	FD with controls	FE with controls	FD with controls	FE with controls	FD with controls	FE with controls
Contemp. coefficient estimate	-0.05 ***	-0.01 **	-0.03	-0.02 ***	-0.03	-0.002
First lag coefficient estimate	-0.05 ***		-0.02		-0.06 **	

By comparing the slope coefficient estimates we received similar results in magnitude, but there are visible differences between the groups and the original panel. The FE coefficient for low-income countries remained statistically significant and turned out to be larger, than for the full panel. The FD estimate for both groups is smaller and statistically not significant, being –a.a3 for both groups. The coefficient of the first lag became slightly larger and remained statistically significant at the 1% level for high-income countries, while it became smaller than for the full panel for low-income countries. We can conclude from the results that some unobserved heterogenity is still present in the previous FD and FE estimates of the full panel when controlling for the above mentioned variables, however, the low- and high-income groups faced a very similar effect in magnitude, therefore the previous estimates can be considered to be robust.

APPENDIX *VII.* | Code

```
rm(list = ls())
library(WDI)
library(data.table)
library(stringr)
library(readr)
library(dplyr)
library(tidyr)
library(ggplot2)
library(gridExtra)
library(plm)
library(stargazer)
library(fBasics)
library(pander)
library(knitr)
library(dplyr)
library(sandwich)
library(lmtest)

setwd("C:/Users/Keresztesi Luca/Box Sync/CEU 2nd trimester/Data Analysis 3/Term Project")

# DOWNLOADING DATA THROUGH World Bank API--------------------------------------------------------

# SEARCHING FOR DATA: Wage and salaried workers, female (% of female employed)
wag_inds <- WDIsearch('wage')
wagCode <- wag_inds[match
                    ("Wage and salaried workers, female (% of females employed)",
                     wag_inds[,2],1)]
# DATA DOWNLOAD: Wage and salaried workers, female (% of females employed)
dat_wag = WDI(
  indicator = wagCode,
  start = 1995, end = 2014)
# FILTERING OUT REGIONS
dt_wag <- data.table(dat_wag)
exclusionList <- dt_wag[,.(itemCnt = .N),by = .(code = dt_wag$iso2c)][1:47, 1]
wagData <- subset(dt_wag, !(dt_wag$iso2c %in%  exclusionList$code))

# SEARCHING FOR DATA: Maternal mortality ratio (per 100,000 live births)
mat_inds <- WDIsearch('maternal')
matCode <- mat_inds[match
                    ("Maternal mortality ratio (modeled estimate, per 100,000 live births)",
                     mat_inds[,2],1)]
# DATA DOWNLOAD: Maternal mortality ratio (per 100,000 live births)
dat_mat = WDI(
  indicator = matCode,
  start = 1995, end = 2014)
# FILTERING OUT REGIONS
dt_mat <- data.table(dat_mat)
exclusionList <- dt_mat[,.(itemCnt = .N),by = .(code = dt_mat$iso2c)][1:47, 1]
matData <- subset(dt_mat, !(dt_mat$iso2c %in%  exclusionList$code))

# SEARCHING FOR DATA: Health expenditure, total (% of GDP)
health_inds <- WDIsearch('health')
healthCode <- health_inds[match
                    ("Health expenditure, total (% of GDP)",
                     health_inds[,2],1)]
# DATA DOWNLOAD: Health expenditure, total (% of GDP)
dat_health = WDI(
  indicator = healthCode,
  start = 1995, end = 2014)
# FILTERING OUT REGIONS
dt_health <- data.table(dat_health)
exclusionList <- dt_health[,.(itemCnt = .N),by = .(code = dt_health$iso2c)][1:47, 1]
healthData <- subset(dt_health, !(dt_health$iso2c %in%  exclusionList$code))

# SEARCHING FOR DATA: GDP per capita
gdp_inds <- WDIsearch('gdp')
grep("2005", gdp_inds, value = TRUE )
gdpppCode <- gdp_inds[match("GDP per capita, PPP (constant 2005 international $)",gdp_inds[,2]),1]
# DATA DOWNLOAD: GDP per capita
dat = WDI(
  indicator=gdpppCode,
  start = 1990, end = 2014)
# FILTERING OUT REGIONS
dt <- data.table(dat)
exclusionList <-dt[,.(itemCnt=.N),by=.(code = dt$iso2c)][1:47,1]
gdpData <- subset(dt, !(dt$iso2c %in%  exclusionList$code))


# SEARCHING FOR DATA: Population, total
pop_inds <- WDIsearch('population')
popCode <- pop_inds[match
                    ("Population, total",
                     pop_inds[,2],1)]
# DATA DOWNLOAD: Population
dat_pop = WDI(
```

```
  indicator = popCode,
  start = 1990, end = 2014)
# FILTERING OUT REGIONS
dt_pop <- data.table(dat_pop)
exclusionList <- dt_pop[,.(itemCnt = .N),by = .(code = dt_pop$iso2c)][1:47, 1]
popData <- subset(dt_pop, !(dt_pop$iso2c %in%  exclusionList$code))

# MERGING DATA INTO PANEL
panelData_1 <- merge(wagData, gdpData,
                     by.x = c("iso2c", "year","country"),
                     by.y = c("iso2c", "year", "country") )
panelData_2 <- merge(healthData, matData,
                     by.x = c("iso2c", "year","country"),
                     by.y = c("iso2c", "year", "country") )
panelData_3 <- merge(popData, panelData_1,
                     by.x = c("iso2c", "year","country"),
                     by.y = c("iso2c", "year", "country") )
# panelData_4 <- merge(highData, mmortData,
#                      by.x = c("iso2c", "year","country"),
#                      by.y = c("iso2c", "year", "country") )
# panelData_5 <- merge(panelData_1, panelData_2,
#                      by.x = c("iso2c", "year","country"),
#                      by.y = c("iso2c", "year", "country") )
# panelData_6 <- merge(panelData_3, panelData_4,
#                      by.x = c("iso2c", "year","country"),
#                      by.y = c("iso2c", "year", "country") )
panelData <- merge(panelData_2, panelData_3,
                   by.x = c("iso2c", "year","country"),
                   by.y = c("iso2c", "year", "country") )

# CLEANING AND TRANSOFMRING THE DATA------------------------------------------------------------

dat <- NULL
dt_wag <- NULL
panelData$iso2c <- NULL

up <- data.table(panelData)
names(up) <- c('year', 'country', 'health', 'mmort', 'pop', 'salaried', 'gdppc')
up$pop <- up$pop/10^6 # Count the population in million
up <- subset(up, up$year < 2015 & up$year > 1994)


# DATA EXPLORATION------------------------------------------------------------------------------

# non-missing observations for each country
bp <- subset(up,
             !is.na(up$salaried) &
               !is.na(up$mmort) &
               !is.na(up$pop) &
               !is.na(up$gdppc) &
               !is.na(up$health)
             )

bp %>%
  count(country) %>%
  arrange(n) %>%
  print(n = 600)

# checking world trends
world_trend <- up %>%
  group_by(year) %>%
  summarise(
    salaried = weighted.mean(salaried, na.rm = TRUE),
    mmort = weighted.mean(mmort, w = pop, na.rm = TRUE),
    health = weighted.mean(health, w = pop, na.rm = TRUE),
    gdppc = weighted.mean(gdppc, w = pop, na.rm = TRUE))

world_trend$rellnsalaried = log(world_trend$salaried) - first(log(world_trend$salaried))
world_trend$rellnhealth = log(world_trend$health) - first(log(world_trend$health))
world_trend$rellngdp = log(world_trend$gdppc) - first(log(world_trend$gdppc))
world_trend$rellnmmort = log(world_trend$mmort) - first(log(world_trend$mmort))

# subplot (1)
p1 <- world_trend %>%
  ggplot(aes(year, mmort)) +
  geom_line(size = 1, color = 'darkgreen') +
  ylab('Maternal mortality ratio (per 100,000 live births)') # y-axis label

# subplot (2)
p2 <- world_trend %>%
  ggplot(aes(x = year)) +
  geom_line(aes(y = rellnmmort), size = 1, linetype = 'dotted', color = 'darkgreen') +
  geom_line(aes(y = rellnsalaried), size = 1, linetype = 'dashed', color = 'firebrick') +
  geom_text(x = 2004, y = -0.25, label = 'Maternal mortality ratio (per 100,000 live births)', color = 'darkgreen') +
  geom_text(x = 2009, y = -0.05, label = 'salaried female (% of total emolyed)', color = 'firebrick') +
```

```
  ylab('log change from 1995')

# arrange these two subplots
grid.arrange(p1, p2)

# subplot (3)
p3 <- world_trend %>%
  ggplot(aes(year, salaried)) +
  geom_line(size = 1, color = 'firebrick') +
  ylab('salaried female (% of total emolyed)') # y-axis label

# subplot (4)
p4 <- world_trend %>%
  ggplot(aes(x = year)) +
  geom_line(aes(y = rellnmmort), size = 1, linetype = 'dotted', color = 'darkgreen') +
  geom_line(aes(y = rellnhealth), size = 1, linetype = 'dashed', color = 'deepskyblue3') +
  geom_text(x = 2004, y = -0.25, label = 'Maternal mortality ratio (per 100,000 live births)', color = 'darkgreen') +
  geom_text(x = 2006, y = 0.01, label = 'Healthcare expenditure (% of GDP)', color = 'deepskyblue3') +
  ylab('log change from 1995')

# arrange these two subplots
grid.arrange(p3, p4)

# country-specific trends: selected countries, difference between 1995 and 2014
interesting_countries <- c(
  "United Kingdom", "Turkey", "Malaysia", "Hungary",
  "Pakistan", "venezuela", "El Salvador", "Macedonia, FYR", "Mexico"
)

up_int <- subset(up, up$country %in% interesting_countries, na.rm = TRUE)
p5 <- up_int %>%
  group_by(country) %>%
  ggplot(aes(year, salaried)) +
  geom_line(aes(color = country, linetype = country), size = 1) +
  ylab('Ratio of wage and salaried of female employment')

p6 <- up_int %>%
  group_by(country) %>%
  ggplot(aes(year, mmort)) +
  geom_line(aes(color = country, linetype = country), size = 1) +
  ylab('Maternal mortality ratio (per 100,000 live births)')

p7 <- up_int %>%
  group_by(country) %>%
  ggplot(aes(year, health)) +
  geom_line(aes(color = country, linetype = country), size = 1) +
  ylab('Health expenditure (% of GDP)')

# arrange these three subplots
grid.arrange(p5, p6, p7)


# examining the balanced panel with no missing values on key variables

ggplot(bp) + aes(x = mmort) +
  geom_histogram (binwidth = 20,
                  fill ='darkgreen') +
  labs(
    x ="Maternal mortality ratio (per 100,000 live births)",
    title ="Maternal mortality ratio (per 100,000 live births) in years 1995-2014") +
  theme_bw()

ggplot(bp) + aes(x = salaried) +
  geom_histogram (binwidth = 1,
                  fill ='firebrick') +
  labs(
    x ="Ratio of wage and salaried of female employment')",
    title ="Ratio of wage and salaried of female employment') in years 1995-2014") +
  theme_bw()


ggplot(bp) + aes(x = health) +
  geom_histogram (binwidth = 0.5,
                  fill ='deepskyblue3') +
  labs(
    x ="Health expenditure (% of GDP)",
    title ="Health expenditure (% of GDP) in years 1995-2014") +
  theme_bw()

ggplot(bp) + aes(x = gdppc) +
  geom_histogram(binwidth = 1000,
                 fill ='grey') +
  labs(
    x = "GDP per capita distribution across all years, constant 2005 prices",
    title = "Histogram of GDP per capita distribution in years 1995-2014") +
```

```
  theme_bw()

stargazer(
  bp,
  header = FALSE,
  out = "summarystats.html",
  type = 'latex',
  omit = 'year',
  title = "Descriptive Statistics for the variables in the balanced panel")

up$lngdppc <- log(up$gdppc)
up$lnmmort <- log(up$mmort)
up$lnpop <- log(up$pop)
up$lnhealth <- log(up$health)


miss_lnmmort <- merge(up[, mean(mmort, na.rm = TRUE), by = country],
                  up[is.na(mmort), .N, by = country], by = "country",
                  all = TRUE)

miss_salaried <- merge(up[, mean(salaried, na.rm = TRUE), by = country],
                    up[is.na(salaried), .N, by = country], by = "country",
                    all = TRUE)

miss_lngdppc <- merge(up[, mean(lngdppc, na.rm = TRUE), by = country],
                      up[is.na(lngdppc), .N, by = country], by = "country",
                      all = TRUE)

miss_health <- merge(up[, mean(health, na.rm = TRUE), by = country],
                     up[is.na(health), .N, by = country], by = "country",
                     all = TRUE)

ggplot(miss_lnmmort) + aes(x = N) +
  geom_histogram (binwidth = 0.1,
                fill ='darkgreen') +
  labs(
    x ="Average number of missing observations",
    title ="Dist. of avg. missing obs. of Log Maternal mortality") +
  theme_bw()

ggplot(miss_salaried) + aes(x = N) +
  geom_histogram (binwidth = 1,
                fill ='firebrick') +
  labs(
    x ="Average number of missing observations",
    title ="Dist. of avg. missing obs. of Ratio of salaried female workers") +
  theme_bw()

ggplot(miss_lngdppc) + aes(x = N) +
  geom_histogram (binwidth = 1,
                fill ='grey') +
  labs(
    x ="Average number of missing observations",
    title ="Dist. of avg. missing obs. of Log GDP per capita") +
  theme_bw()

ggplot(miss_health) + aes(x = N) +
  geom_histogram (binwidth = 1,
                fill ='deepskyblue3') +
  labs(
    x ="Average number of missing observations",
    title ="Dist. of avg. missing obs. of Healthcare expenditure (% of GDP)") +
  theme_bw()

ggplot(miss_lnmmort, aes(y = V1, x = N)) +
  geom_point(color = "darkgreen") +
  geom_smooth(method = "lm", se = FALSE) +
  labs(
    x ="Average number of missing observations",
    y = "Mean Log Maternal Mortality",
    title ="Mean Log Maternal mortality in years 1995-2014") +
  theme_bw()

ggplot(miss_salaried, aes(y = V1, x = N)) +
  geom_point(color = "firebrick") +
  geom_smooth(method = "lm", se = FALSE) +
  labs(
    x ="Average number of missing observations",
    y = "Mean Ratio of salaried female workers",
    title ="Mean Ratio of salaried female workers in years 1995-2014") +
  theme_bw()

ggplot(miss_lngdppc, aes(y = V1, x = N)) +
  geom_point(color = "grey") +
  geom_smooth(method = "lm", se = FALSE) +
```

```
  labs(
    x ="Average number of missing observations",
    y = "Mean Log GDP per capita",
    title ="Mean Log GDP per capita in years 1995-2014") +
  theme_bw()

ggplot(miss_health, aes(y = V1, x = N)) +
  geom_point(color = "deepskyblue3") +
  geom_smooth(method = "lm", se = FALSE) +
  labs(
    x ="Average number of missing observations",
    y = "Mean Health expenditure (% of GDP)",
    title ="Mean Health expenditure (% of GDP) in years 1995-2014") +
  theme_bw()

# ANALYSIS-----------------------------------------------------------------------------------

# xcountry OLS: pooled OLS regardless country difference
dt <- up[, .(
  m_salaried = mean(salaried, na.rm = TRUE),
  m_lnmmort= mean(lnmmort, na.rm = TRUE),
  m_lngdppc = mean(lngdppc, na.rm = TRUE),
  m_lnpop = mean(lnpop, na.rm = TRUE),
  m_lnhealth = mean(lnhealth, na.rm = TRUE)),
  by = .(country)]

p8 <- ggplot(dt, aes(x = m_salaried, y = m_lnmmort)) +
  geom_point(size = 1.5, aes(col = (m_lnhealth))) +
  geom_smooth() +
  labs(
    x = "Average salaried female workers (% of total) per country",
    y = "Average Log Maternal Mortality per country",
    title ="Average Log Maternal Mortality on salaried female workers (% of total)") +
  scale_color_distiller("Log Health expenditure (% of GDP)", palette = "Spectral") +
  theme_bw()

p9 <- ggplot(dt, aes(x = m_salaried, y = m_lnmmort)) +
  geom_point(size = 1.5, aes(col = (m_lngdppc))) +
  geom_smooth() +
  labs(
    x = "Average salaried female workers (% of total) per country",
    y = "Average Log Maternal Mortality per country",
    title ="Average Log Maternal Mortality on salaried female workers (% of total)") +
  scale_color_distiller("Log GDP per capita (const. 2005 $)", palette = "Spectral") +
  theme_bw()

grid.arrange(p8, p9)


# OLS regressions
ols1995 <- lm(data = up[year == 1995], lnmmort ~ salaried)
ols2007 <- lm(data = up[year == 2007], lnmmort ~ salaried)
ols2014 <- lm(data = up[year == 2014], lnmmort ~ salaried)
ols2007_c <- lm(data = up[year == 2007], lnmmort ~ salaried + lnhealth + lngdppc)


ols_models <- list(ols1995, ols2007, ols2014, ols2007_c)

cov_ols_1         <- vcovHC(ols1995, type = "HC1")
rob_ols_1         <- sqrt(diag(cov_ols_1))
cov_ols_2         <- vcovHC(ols2007, type = "HC1")
rob_ols_2         <- sqrt(diag(cov_ols_2))
cov_ols_3         <- vcovHC(ols2014, type = "HC1")
rob_ols_3         <- sqrt(diag(cov_ols_3))
cov_ols_4         <- vcovHC(ols2007_c, type = "HC1")
rob_ols_4         <- sqrt(diag(cov_ols_4))

stargazer(
  title = "OLS models",
  list(ols1995, ols2007, ols2014, ols2007_c), digits = 2,
  column.labels = c('ols1995', 'ols2007', 'ols2014', 'ols2007c'),
  model.names = FALSE,
  omit.stat = c("adj.rsq", "f"),
  dep.var.caption = 'Dependent variable: Maternal Mortality',
  out = "OLS.html",
  notes.align = "l",
  se = list(rob_ols_1, rob_ols_2, rob_ols_3, rob_ols_4),
  header = FALSE,
  type ='latex'
)


# FE with explicit time dummies

fe1 <- plm(
```

```
  lnmmort ~ salaried + year, data = up,
  model = 'within')

fe2 <- plm(
  lnmmort ~ salaried + lngdppc + year, data = up,
  model = 'within')

fe3 <- plm(
  lnmmort ~ salaried + lnhealth + year, data = up,
  model = 'within')

fe4 <- plm(
  lnmmort ~ salaried + lngdppc + lnhealth + lnpop + year, data = up,
  model = 'within')

fe_models <- list(fe1, fe2, fe3, fe4)

cov_fe_1        <- vcovSCC(fe1, type = "HC1")
rob_fe_1        <- sqrt(diag(cov_fe_1))
cov_fe_2        <- vcovSCC(fe2, type = "HC1")
rob_fe_2        <- sqrt(diag(cov_fe_2))
cov_fe_3        <- vcovSCC(fe3, type = "HC1")
rob_fe_3        <- sqrt(diag(cov_fe_3))
cov_fe_4        <- vcovSCC(fe4, type = "HC1")
rob_fe_4        <- sqrt(diag(cov_fe_4))

stargazer(
  title = "Comparing FE models with expicit time dummies",
  list(fe_models), digits = 2,
  column.labels = c('FE1', 'FE2', 'FE3', 'FE4'),
  model.names = FALSE,
  omit.stat = c("adj.rsq", "f", "ser"),
  dep.var.caption = 'Dependent variable: Log Maternal Mortality',
  out = "FE.html",
  notes.align = "l",
  se = list(rob_fe_1, rob_fe_2, rob_fe_3, rob_fe_4),
  dep.var.labels.include = FALSE,
  header = FALSE,
  type ='latex'
)


# FD with explicit time dummies and lags

diff1 <- plm(
  diff(lnmmort) ~ diff(salaried) + year,
  data = up, model = 'pooling'
)

diff2 <- plm(
  diff(lnmmort) ~ diff(salaried) + stats::lag(diff(salaried), 1:2) + year,
  data = up, model = 'pooling'
)

diff3 <- plm(
  diff(lnmmort) ~ diff(salaried) + stats::lag(diff(salaried), 1:4) + year,
  data = up, model = 'pooling'
)

diff4 <- plm(
  diff(lnmmort) ~ diff(salaried) + stats::lag(diff(salaried), 1:6) + year,
  data = up, model = 'pooling'
)

cov_fd_1        <- vcovSCC(diff1, type = "HC1")
rob_fd_1    <- sqrt(diag(cov_fd_1))
cov_fd_2        <- vcovSCC(diff2, type = "HC1")
rob_fd_2    <- sqrt(diag(cov_fd_2))
cov_fd_3        <- vcovSCC(diff3, type = "HC1")
rob_fd_3    <- sqrt(diag(cov_fd_3))
cov_fd_4        <- vcovSCC(diff4, type = "HC1")
rob_fd_4    <- sqrt(diag(cov_fd_4))

fd_models <- list(diff1, diff2, diff3, diff4)

stargazer(
  title = "Comparing FD models with expicit time dummies",
  list(fd_models), digits = 2,
  column.labels = c('FD1', 'FD2', 'FD3', 'FD4'),
  model.names = FALSE,
  omit.stat = c("adj.rsq", "f", "ser"),
  dep.var.caption = 'Dependent variable: Log Maternal Mortality',
  out = "FD.html",
  notes.align = "l",
  se = list(rob_fd_1, rob_fd_2, rob_fd_3, rob_fd_4),
```

```
  add.lines = list(
    c("Cumulative Coeff", '-0.04', '-0.045', '-0.04', '1.26')),
  dep.var.labels.include = FALSE,
  header = FALSE,
  type ='latex'
)

# FD with explicit time dummies, lags and controls

diff_c1 <- plm(
  diff(lnmmort) ~ diff(salaried) + year + diff(lnhealth) + diff(lngdppc) + diff(pop),
  data = up, model = 'pooling'
)

diff_c2 <- plm(
  diff(lnmmort) ~ diff(salaried) + stats::lag(diff(salaried), 1:2) +
    stats::lag(diff(lnhealth), 1:2) +
    stats::lag(diff(lngdppc), 1:2) +
    stats::lag(diff(lnpop), 1:2) +
    year,
  data = up, model = 'pooling'
)

diff_c3 <- plm(
  diff(lnmmort) ~ diff(salaried) + stats::lag(diff(salaried), 1:4) +
    stats::lag(diff(lnhealth), 1:4) +
    stats::lag(diff(lngdppc), 1:4) +
    stats::lag(diff(lnpop), 1:4) +
    year,
  data = up, model = 'pooling'
)


cov_fdc_1         <- vcovSCC(diff_c1, type = "HC1")
rob_fdc_1    <- sqrt(diag(cov_fdc_1))
cov_fdc_2         <- vcovSCC(diff_c2, type = "HC1")
rob_fdc_2    <- sqrt(diag(cov_fdc_2))
cov_fdc_3         <- vcovSCC(diff_c3, type = "HC1")
rob_fdc_3    <- sqrt(diag(cov_fdc_3))

fd_models_controls <- list(diff_c1, diff_c2, diff_c3)

stargazer(
  title = "Comparing FD models with expicit time dummies and controls",
  list(fd_models_controls), digits = 2,
  column.labels = c('FD5', 'FD6', 'FD7'),
  model.names = FALSE,
  omit.stat = c("adj.rsq", "f", "ser"),
  dep.var.caption = 'Dependent variable: Log Maternal Mortality',
  out = "FD_controls.html",
  notes.align = "l",
  se = list(rob_fdc_1, rob_fdc_2, rob_fdc_3),
  add.lines = list(
    c("Cumulative Coeff", '-0.01', '-0.104', '0.02')),
  dep.var.labels.include = FALSE,
  header = FALSE,
  type ='latex'
)

# Multiple models

multiple_models = list(ols2007_c, fe4, diff2, diff_c2)

stargazer(
  title = "Comparing multiple models",
  list(multiple_models), digits = 2,
  column.labels = c('OLS2007c', 'FE4c', 'FD2', 'FD2c'),
  model.names = FALSE,
  omit.stat = c("adj.rsq", "f", "ser"),
  dep.var.caption = 'Dependent variable: Log Maternal mortality',
  out = "Multiple.html",
  notes.align = "l",
  se = list(rob_ols_4, rob_fe_4, rob_fd_2, rob_fdc_2),
  dep.var.labels.include = FALSE,
  header = FALSE,
  type ='latex',
  omit = c('year', 'lngdppc', 'lnhealth', 'lnpop',
           'stats::lag(diff(lnhealth), 1:2)',
           'stats::lag(diff(lngdppc), 1:2)',
           'stats::lag(diff(lnpop), 1:2)')
)

# Grouping countries by income

# Group 1
```

```
up %>%
  select(salaried, gdppc, pop) %>%
  as.data.frame() %>%
  stargazer(
    type = 'text', flip = TRUE, digits = 1,
    summary.stat = c('mean', 'min', 'median', 'p25', 'p75', 'max', 'n')
  )

up$group <- ifelse(up$gdppc <= 8500,1,2)
up_1 <- subset(up, up$group == 1)
up_2 <- subset(up, up$group == 2)

g_fd2c <- plm(
  diff(lnmmort) ~ diff(salaried) + stats::lag(diff(salaried), 1:2) +
    stats::lag(diff(lnhealth), 1:2) +
    stats::lag(diff(lngdppc), 1:2) +
    stats::lag(diff(lnpop), 1:2) +
    year,
  data = up_1, model = 'pooling'
)

g_fe4c <- plm(
  lnmmort ~ salaried + lngdppc + lnhealth + lnpop + year,
  data = up_1,
  model = 'within')

# Group 2

gg_fd2c <- plm(
  diff(lnmmort) ~ diff(salaried) + stats::lag(diff(salaried), 1:2) +
    stats::lag(diff(lnhealth), 1:2) +
    stats::lag(diff(lngdppc), 1:2) +
    stats::lag(diff(lnpop), 1:2) +
    year,
  data = up_2, model = 'pooling'
)

gg_fe4c <- plm(
  lnmmort ~ salaried + lngdppc + lnhealth + lnpop + year,
  data = up_2,
  model = 'within')

# Standard errors were adjusted to make sure robust standard errors are used
g_cov_fd        <- vcovSCC(g_fd2c, type = "HC1")
g_rob_fd        <- sqrt(diag(g_cov_fd))
g_cov_fe        <- vcovSCC(g_fe4c, type = "HC1")
g_rob_fe        <- sqrt(diag(g_cov_fe))
gg_cov_fd        <- vcovSCC(gg_fd2c, type = "HC1")
gg_rob_fd        <- sqrt(diag(gg_cov_fd))
gg_cov_fe        <- vcovSCC(gg_fe4c, type = "HC1")
gg_rob_fe        <- sqrt(diag(gg_cov_fe))

stargazer(
  title = "Comparing multiple models for high- and low-income countries",
  list(g_fd2c, g_fe4c, gg_fd2c, gg_fe4c), digits = 2,
  column.labels = c('FD2c, G:1', 'FE4c, G:1', 'FD2c, G:2', 'FE4c, G:2'),
  model.names = FALSE,
  omit.stat = c("adj.rsq", "f", "ser"),
  dep.var.caption = 'Dependent variable: Log Maternal mortality',
  out = "Grouping.html",
  notes.align = "l",
  se = list(g_rob_fd, g_rob_fe, gg_rob_fd, gg_rob_fe),
  dep.var.labels.include = FALSE,
  header = FALSE,
  type ='latex',
  omit = c('year', 'lngdppc', 'lnhealth', 'lnpop',
           'stats::lag(diff(lnhealth), 1:2)',
           'stats::lag(diff(lngdppc), 1:2)',
           'stats::lag(diff(lnpop), 1:2)')
)
```