Topics in Applied Microeconomics
Referee Report

Brian Wang

January 15, 2021

Title: Arrival of Young Talent: The Send-Down Movement and Rural Education in China

Authors: Yi Chen, Ziying Fan, Xiaomin Gu, and Li-An Zhou

Source: American Economic Review Vol. 110, NO. 11, 2020: 3393-3430

Abstract

Using cohort DID method, this paper estimates the effects of send-down youth on rural education. In this report, I start with an analysis of the data and contributions of this paper. I select the cohort DID results for replication. In order to show the robustness of the main result, I first replicate the by-cohort DID to test the key parallel-trend assumption. I proceed to conduct a minor extension which controls one more personal attribute, and I show the main result remains the same. Lastly, I discuss the weakness of the paper, and extend my report by proposing a DID model with linear constraints. This identifies the partial exposure effects for each cohort with a specific functional form, and the results support both the original paper and my extension model.

1 Summary of this Paper

This paper looks into the effect of send-down youth on rural education. During 1968 to the late 1970s, China's central government launched the "send-down movement", which mandated about 16 million urban youth (namely SDY, send-down youth) to temporarily resettle in rural areas. This provides a unique chance for estimating the causal effect: first, the treatment was implemented in a top-down, mandatory manner, thus preventing self-selection. Secondly, migration was highly restricted during the treatment period, which limits attrition bias. Thirdly, the treatment is temporary. By the time rural children grew up, most SDYs had left the rural area. They will not be counted as rural residents, therefore keeping the treatment group uncontaminated.

The **data** used in this research is extremely precious: together with individual-level population census (1982,1990, and 2000), a hand-collected dataset from over 3000 book-length local gazetteers was constructed to calculate number of SDYs at county level. Gazetteers are local record of big events. Being seen as a regional glory, they are usually kept in good conditions with detailed records. In terms of number of SDYs, these records could undoubtedly be regraded as the most reliable source of information.

The paper adopts a difference-in difference **methodology.** I select the **cohort-DID model** as the main focus of this report:

$$\begin{aligned} Y\_Edu_{i,g,c,p} = \beta_0 + {\color{red}\beta_1}\%SDY_{c,p} \times I(1956 \leqslant g \leqslant 1969) \\ + \beta_2 \mathbf{X}_{i,g,c,p} + \lambda_c + \mu_{g,p} + \Lambda_c \times \mu_g + \epsilon_{i,g,c,p} \end{aligned} \tag{1}$$

The key **assumption for identification** is the parallel-trend assumption. This is supported by several pieces of evidence. First, they show that there was no pre-existing heterogeneous trends with a by-cohort DID regression[1]. Second, possible heterogeneity due to local government's effort to expand rural education is examined. After controlling for number of local schools, it is shown to be unimportant for the size of β_1. Thirdly, other factors that could cause different trends (such as different grain production level) are examined in the robustness test and shown to have minor influence on the main result[2].

The **main results** suggest that exposure to SDYs increased rural children's education by at

[1] see replication in section 3

[2] not replicated as it will need another table to report

least 0.072 years. In other words, the SDYs resulted in a 17.6 million person-year increase in rural children's schooling[3]. The result remains robust against changes in bandwidths of treated cohort, changes in the denominator calculating the density of SDYs, and exclusion (inclusion) of several provinces that are more (less) impacted by SDYs. Moreover, they conducted subgroup analysis, and showed that the SDYs' effect is about twice as large for girls than boys. Also, the effect is greater on less-educated groups[4].

The **mechanism** of such impact is probed through a mediation analysis. The paper showed that that better-educated SDYs contributed to the local pool of teachers, and hence increases rural education level.

Overall, this paper has an extensive cover with complete discussions. I chose to replicate the cohort-DID result **because** it examines the main causal effect, and is the basis for probing mechanisms. The next section will provide an analysis in this paper's strength and weakness. In section 3 we will replicate the main result. In section 4 we proposal an extension on the identification methods using a specific functional form. We conclude the report in section 5.

2 Contributions, Strength and Weakness

This paper contributes to the literature of education in developing countries in examining the individual-level outcomes of an increase in teacher supply. Other supply side factors have been studied, for example, by Miguel and Kremer 2004 which focuses on child health. This paper shows an increase of qualified teacher supply has a persistent effect on likelihood of going to senior high school, attitude towards education and even occupational choices.

This paper has useful policy implications. As the paper is the first of its kind to systematically estimate the impact of the send-down movement, the result can be insightful for future education policies. In specific, this could encourage new policy designed nowadays which encourage university graduates to support education in mountain areas for several years before they leave the area for further studies.

[3] see replication in section 3

[4] not replicated as it will need another table to report

Another strength of this paper is that they use a proper identification strategy to address the research question. The authors have a good understanding of cohort DID and by-cohort DID (Duflo 2001), and are clearly aware of the importance of the parallel-trend assumption. In particular, they apply state-of-art regression methods to control for the potentially heterogeneous trends by allowing high dimensional fixed effects (Correia 2017). Their interpretation of the parameter is also appropriate in that they explicitly calculate the intensity of treatment, instead of a treatment dummy, and calculated the average impact of SDYs based on average density of SDYs. Finally, they properly controlled for other potential mediations, such as heterogeneous increase of crop yield. This provides a solid support for their conclusion that the main mechanism through which the causal effect takes place is the increase of qualified teacher supply.

One weakness of this paper is that the cohort DID, which gives the main result of the paper, estimates only the average treatment effect—it could be further studied what the treatment effects on different cohorts or individuals are. For example, we could expect that individuals who were born in an earlier treatment year were exposed to SDYs for shorter time, thus have a lower coefficient compared with individuals born in later treatment year. Following this motivation, I develop my extension in section 4.

3 Main Results Replication and Comparison

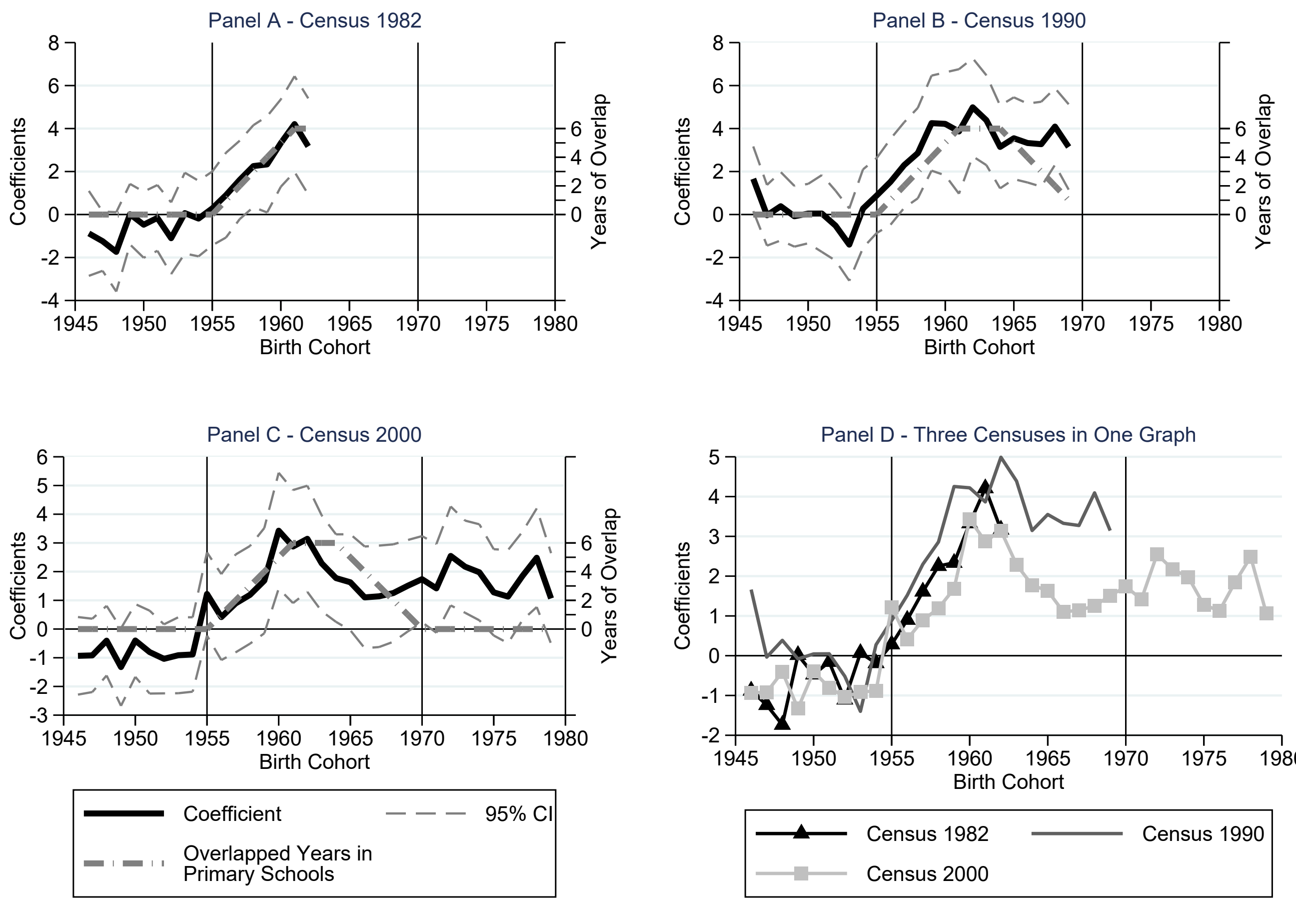
3.1 Identification Assumption (By-Cohort DID)

The key assumption for identifying the DID model is the parallel-trend assumption. In specific, each county's trend in eduation should not be related to the intensity of treatment (density of SDYs). By running the following regression, the paper supports the assumption:

$$\begin{aligned} Y\_Edu_{i,g,c,p} = \beta_0 + \sum_{\gamma=1946}^{1969} \beta_{1,\gamma} \%SDY_{c,p} \times I(g=\gamma) \\ + \beta_2 \mathbf{X}_{i,g,c,p} + \lambda_c + \mu_{g,p} + \Lambda_c \times \mu_g + \epsilon_{i,g,c,p} \end{aligned} \tag{2}$$

in which the parameter of interest is β_1, which captures the treatment effect per density of SDY ($\%SDY_{c,p}$) for each year. We control for time ($\mu_{g,p}$) and county fixed effects (λ_c), a cross term between them ($\Lambda_c \times \mu_g$), and several individual attributes ($\mathbf{X}_{i,g,c,p}$).

Figure 1: By-Cohort DID (Figure 3 in the paper)



My replication is presented in figure 1. With the help of high dimensional fixed effect model (Correia 2017), I estimated all 24 ($= 1969 - 1946 + 1$) coefficients efficiently. Panel A to C applies different census data to equation 2. Each dot in each panel represents a $\beta_{1,\gamma}$ and the line sketches the treatment effect of different years. The replication resembles Figure 3 in the paper perfectly.

As the coefficients fluctuate around zero before 1956, it supports the assumption that there were no heterogeneous cohort trends at least prior to the treatment. In addition, we can extract more knowledge from figure 1. Exposure to SDYs has an accumulative effect during the starting years (1956-1961), maintained around 3 at its plateaux, and diminishes from 1963 onwards. This is especially shown in Panel C, where longer observations are available.

3.2 A Minor Extension and Comparison (Cohort DID)

Next I replicate the estimation result for equation 1. The treatment group in the model are rural children born between 1956 and 1969, as they were at the age of primary education when SDY arrived. The control group are rural individuals born between 1946-1955. In the paper, individual attributes ($\mathbf{X}_{i,g,c,p}$) includes **only ethnic group and gender.**

First, I **stick to the original $\mathbf{X}_{i,g,c,p}$**. My replication result is shown **in the appendix (Table 2).** Besides the asterisks I added, it is exactly the same as Table 3 in the paper. Column (1) and (2) uses education year as the measure for Y_Edu, column (3) and (4) uses graduation from primary school (year 1-6), and column (5) and (6) uses graduation from junior high school (year 7-9). We see that the coefficients of intensity of treatment (density of SDYs), after controlling for time and county level fixed effects, are all significantly positive for rural areas, while insignificant for urban areas. Column (7) and (8) conducted two placebo test in which individuals born between 1946 and 1950 (or 1970-1974) are assumed to be a treatment group. The coefficients for them are insignificant, which also supports the parallel-trend assumption.

Secondly, to make one step forward, **I include one more attribute**: $Local\_1985$ in $\mathbf{X}_{i,g,c,p}$, and keep other variables the same as that in the paper. $Local\_1985$ is a 0/1 variable describing the individual is a local resident of county in 1985, available only in 1990 census. This is probably the reason why the authors didn't include it in their analysis.

I control one more personal attribute ($Local\_1985$) to see if the main parameters of interest (effect of SDYs in rural area) will change. The results are shown below **(Table 1)**.

The results are very similar to Table 2. For the coefficient of density of treatment using 1990 census rural sample, we see it decreases from 3.237 to 3.207, maintaining significantly positive at 0.1% level. Similarly, other parameters of interest remain the same. Indeed, the result in the paper has been shown to be robust against various controls, and it is not surprising for us to find Table 1 resembling Table 2.

Table 1: Minor Extension of Cohort-DID

The Effect of SDYs on the Educational Attainment of Rural Children

Dependent Variables	Years of education		Complete primary		Complete junior high		Placebo I (1990)	Placebo II (1990)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample	rural	urban	rural	urban	rural	urban	(1946-1950) versus (1951-1955)	(1970-1974) versus (1975-1979)
Local density of received SDYs × affected cohorts (1956-1969)	3.207***	-0.0453	0.437***	-0.0621	0.763***	-0.0547		
	(0.701)	(0.523)	(0.0871)	(0.0606)	(0.121)	(0.103)		
Local density of received SDYs × affected cohorts (placebo)							-0.817	-0.432
							(0.576)	(0.319)
Male	1.877***	0.654***	0.201***	0.0322***	0.203***	0.0544***	2.286***	0.665***
	(0.0284)	(0.0265)	(0.00362)	(0.00227)	(0.00286)	(0.00314)	(0.0300)	(0.0150)
Han ethnic	0.152***	-0.00576	0.0216***	0.00973*	0.00688	0.0176**	0.0802	0.477***
	(0.0565)	(0.0799)	(0.00769)	(0.00541)	(0.00679)	(0.00872)	(0.0554)	(0.0401)
local_1985	-0.222***	-0.664***	-0.0291***	0.0126*	-0.0317***	-0.00993	-0.00778	
	(0.0274)	(0.115)	(0.00357)	(0.00682)	(0.00420)	(0.0116)	(0.0820)	
Observations	2,775,858	417,883	2,775,858	417,883	2,775,858	417,883	960,123	947,025
R^2	0.293	0.228	0.258	0.106	0.212	0.198	0.267	0.216
$\overline{Y}$	5.372	8.882	0.616	0.911	0.205	0.670		

Notes: Robust standard errors are clustered at county level; *p<0.05, ** p<0.01, *** p<0.001

4 Extension: Partial Exposure

The main results using equation 1 identify the average treatment effect of all individuals born across 1956-1969. Besides the average effect, we are also interested in how the effect change for different cohorts. As we observed from figure 1, the coefficient **first increase, then plateaux, and then decrease.** Could there be a good explanation for this pattern?

One possible reason is that individuals who were born early only received 1 or 2 years of junior high education from the SDYs, and thus the earlier cohorts showed a smaller coefficient. Similarly, for individuals who were born in 1969, they were only age 9 when most SDYs left the rural area (1978), until then they had received only at most 3 years of primary education. Therefore, the coefficients decreases for later cohorts. Moreover, the maximum education delivered by SDYs on any individual will be 9 years (6 years primary + 3 years junior high), which could explain the plateaux.

Borrowing from Bleakley (2010), where they estimates the impact of malaria eradication campaign over a long period, I would like to formalize the above idea with a DID model with constraints capturing the partial exposure effects:

$$\begin{aligned}
Y_Edu_{i,g,c,p} = \beta_0 + &\sum_{\gamma=1946}^{1955} \beta_{1,\gamma} \%SDY_{c,p} \times I_\gamma + \sum_{\gamma=1956}^{1960} [\beta_{1,1956} + (\gamma - 1956) \cdot k_{inc}] \%SDY_{c,p} \times I_\gamma \\
&+ \sum_{\gamma=1961}^{1963} \beta_{1,1961} \%SDY_{c,p} \times I_\gamma + \sum_{\gamma=1964}^{1969} [\beta_{1,1964} - (\gamma - 1956) \cdot k_{dec}] \%SDY_{c,p} \times I_\gamma \\
&+ \sum_{\gamma=1969}^{1979} \beta_{1,\gamma} \%SDY_{c,p} \times I_\gamma + \beta_2 \mathbf{X}_{i,g,c,p} + \lambda_c + \mu_{g,p} + \Lambda_c \times \mu_g + \epsilon_{i,g,c,p}
\end{aligned} \tag{3}$$

in which year 1956-1960 is the increasing period (since student born in 1961 will starting grade 1 when SDYs arrive), 1961-1963 is the plateaux, and 1964-1969 is the decreasing period (since student born after 1964 would not have grade 12 at 1979 under SDYs impact). We only impose linear constraint during treatment period (1956-1969).

To estimate this model, we need to combine constraint regression methods and high dimensional fixed effects models. Unfortunately, *reghdfe* (regression with high dimensional FE) has not been made compatible with *cnsreg* (constrained regression) in STATA. As a compromise, we have to leave out the cross term between time and province ($\Lambda_c \times \mu_g$). Fortunately, the regression results controlling only λ_c and $\mu_{g,p}$ is similar to the original estimates in Table 2 [5]. This allows us to estimate partial exposure effects using *areg* on the following equation:

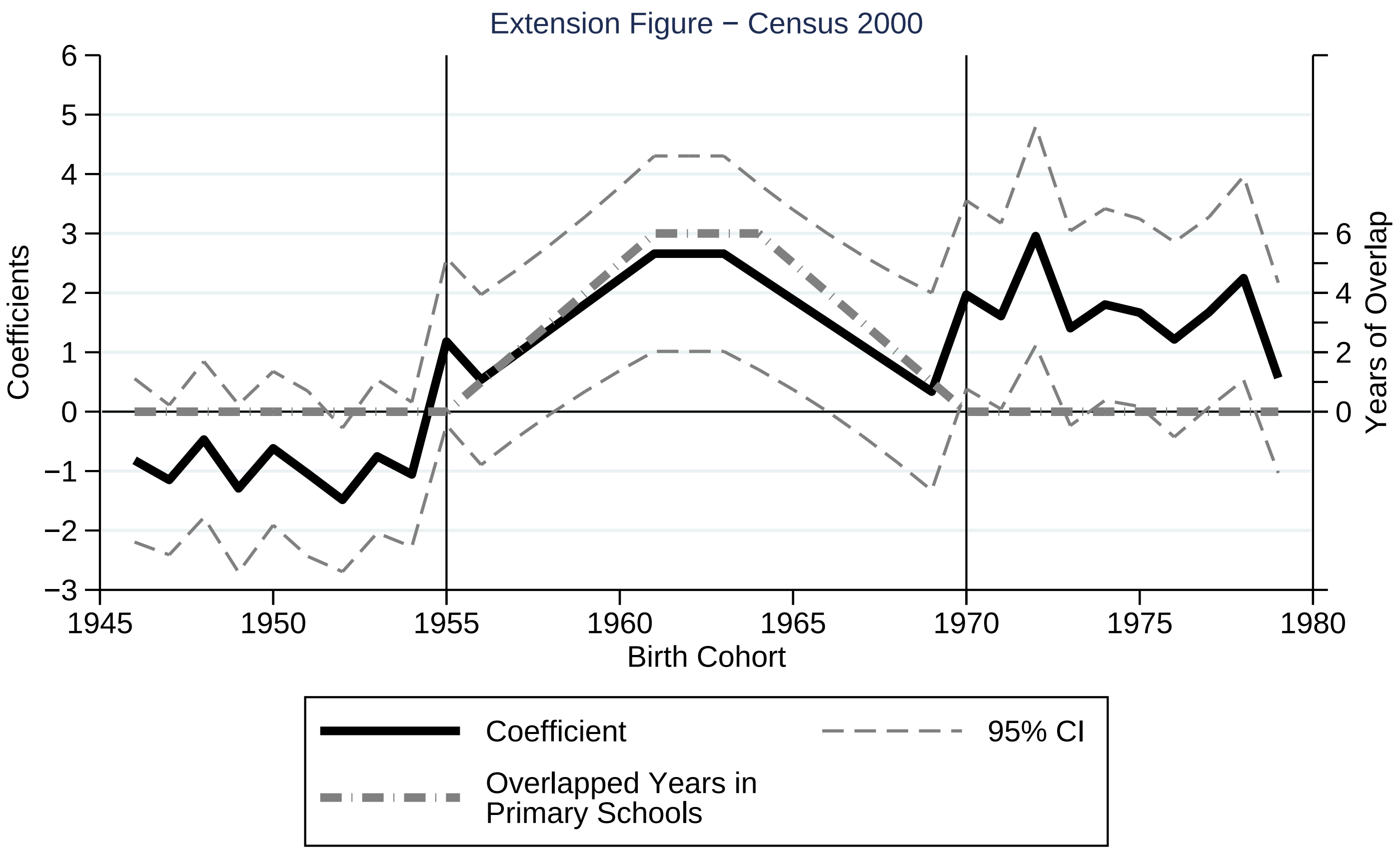
$$Y\_Edu_{i,g,c,p} = \beta_0 + \sum_{\gamma=1946}^{1969} \beta_{1,\gamma} \%SDY_{c,p} \times I(g = \gamma) + \beta_2 \mathbf{X}_{i,g,c,p} + \lambda_c + \mu_{g,p} + \epsilon_{i,g,c,p} \tag{4}$$

in which $\beta_{1,\gamma}{}'s$ satisfy twelve linear constraints specified by equation 3

The results using the above identification strategy is reported concisely in figure 2. The hump-shape in the treatment period (1956-1969) represents the partial exposure effects:

[5] not included in this report as it will require another table

Figure 2: Partial Exposure



Comparing figure 2 with Panel C (which both use data from census 2000) in figure 1, we see that my extension resembles the original result very much. In addition, we estimate $k_{inc} = 0.394$ and $k_{dec} = 0.365$. On one hand, the fact that figure 2 resembles Panel C supports the conclusion in the original paper. We also show that the SDY's impact on education is significant throughout the treatment period (coefficient at around 3) and persists even after 1969. On the other hand, figure 2 also supports our specific assumption that partial exposure takes the form of increase-plateaux-decrease. This supports our intuition that the SDYs' effect on rural education is based on years that treated individuals are born in, and it is the partial exposure that explains the observation in figure 1 better.

5 Conclusion

This paper studies a very meaningful question — how send-down youth improved rural education, and it has been very influential in the research of education in developing countries. The main result is very robust against many specific extensions. Indeed, our replication confirms the robustness against potential control for other personal traits. Therefore, what I did for extension is to propose a specific functional form and check how it explains their data. It is indeed my pleasure to find that a partial effect model could generate a line graph (figure 3) similar to the by-cohort DID regression results (figure 1). It helps to describes the varying effect across cohorts better.

6 Appendix: Replication of Original Table 3 in the Paper

Table 2: Table 3 in the paper

The Effect of SDYs on the Educational Attainment of Rural Children (1990 Census)

Dependent Variables	Years of education		Complete primary		Complete junior high		Placebo I (1990)	Placebo II (1990)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample	rural	urban	rural	urban	rural	urban	(1946-1950) versus (1951-1955)	(1970-1974) versus (1975-1979)
Local density of received SDYs × affected cohorts (1956-1969)	3.237***	0.151	0.441***	-0.0658	0.767***	-0.0517		
	(0.701)	(0.517)	(0.0873)	(0.0611)	(0.121)	(0.103)		
Local density of received SDYs × affected cohorts (placebo)							-0.817	-0.432
							(0.576)	(0.319)
Male	1.874***	0.668***	0.201***	0.0319***	0.203***	0.0546***	2.286***	0.665***
	(0.0284)	(0.0256)	(0.00361)	(0.00227)	(0.00285)	(0.00316)	(0.0300)	(0.0150)
Han ethnic	0.150***	3.34e-05	0.0213***	0.00962*	0.00657	0.0177**	0.0802	0.477***
	(0.0565)	(0.0811)	(0.00769)	(0.00540)	(0.00679)	(0.00875)	(0.0554)	(0.0401)
Observations	2,775,858	417,883	2,775,858	417,883	2,775,858	417,883	960,123	947,025
R^2	0.293	0.225	0.258	0.106	0.212	0.198	0.267	0.216
$\overline{Y}$ of control group	5.372	8.882	0.616	0.911	0.205	0.670		

Robust Standard errors are clustered at the county level; *p<0.05, ** p<0.01, *** p<0.001

References

[1] Joshua D Angrist and Alan B Keueger. Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4):979–1014, 1991.

[2] Hoyt Bleakley. Malaria eradication in the americas: A retrospective analysis of childhood exposure. *American Economic Journal: Applied Economics*, 2(2):1–45, 2010.

[3] Yi Chen, Ziying Fan, Xiaomin Gu, and Li-An Zhou. Arrival of young talent: The send-down movement and rural education in china. *American Economic Review*, 110(11):3393–3430, 2020.

[4] Yi Chen, Sheng Jiang, and Li-An Zhou. Estimating returns to education in urban china: Evidence from a natural experiment in schooling reform. *Journal of Comparative Economics*, 48(1):218–233, 2020.

[5] Sergio Correia. Linear models with high-dimensional fixed effects: An efficient and feasible estimator. Technical report, 2016. Working Paper.

[6] Esther Duflo. Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment. *American economic review*, 91(4):795–813, 2001.

[7] Edward Miguel and Michael Kremer. Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72(1):159–217, 2004.