## Do College Graduates Serving as Village Officials Help Rural China?

#### **ONLINE APPENDIX**

#### By GUOJUN HE AND SHAODA WANG

### Appendix A. CGVO Assignment

In most of China, the assignment of CGVOs is determined entirely by higher levels of government, while villages and CGVOs are not allowed to choose. However, since the exact assignment rules are not known to us, it is important to understand the factors determining CGVO assignment.

There are two main hypotheses regarding assignment decisions. The first is that higher levels of government choose villages based on time-invariant characteristics. For example, governments may prioritize richer and/or larger villages where they expect a CGVO's expertise to help boost economic development. The second hypothesis is that higher levels of government assign CGVOs in response to local economic shocks.

We first test whether the treated villages were systematically different from the control villages before the CGVO program was launched along a variety of socio-economic variables in a cross-sectional setting. We estimate a logit model in which the dependent variable is whether a village has a CGVO during our sample period, and the independent variables are socio-economic conditions in 2006, a year before the CGVO program started to expand.

The regression results are shown in columns (1) to (4) of Appendix Table 1. First, village population and per capita net income are included to test whether CGVO assignments are affected by village size or income. We find no relationship between CGVO assignment and village size or income. Second, we add the outcomes of interest in the regressions, i.e. subsidized population (number of subsidized residents per 1,000 people), poor-quality housing (number of poor-quality houses per 100 households), and registered poor households (number of registered poor households per 100 households). Again, none of them are statistically significant. Third, we include local government size (number of government officials in the

village council) and quality of government officials (proportion of government officials educated to a level of "high school and above") in the regression. The results show that CGVO assignment is uncorrelated with village government size or quality. Finally, a set of time-invariant basic village characteristics are also included, including terrain (flat, hilly or mountainous), its main industry (agriculture, forestry, livestock or fishing), whether the village is located in a suburb, whether it forms a town center, and whether it is a designated poor village. None of them are statistically significant.

An alternative way to test these relationships is to fully exploit the longitudinal structure of the data and estimate the association between CGVO assignment and village-level socioeconomic variables using a logit model with duration dependence. Specifically, the probability of a village receiving a CGVO at time t is modeled as:<sup>1</sup>

(1) 
$$P(CGVO_{it} = 1|X_{it}) = \frac{e^{X_{it}\beta + f(t)}}{1 + e^{X_{it}\beta + f(t)}}$$

where  $CGVO_{it}$  is a dummy variable, which equals 1 if village *i* has a CGVO in year t, and 0 otherwise,  $P(CGVO_{it} = 1|X_{it}) = h(t, X_{it})$  is the probability of receiving a CGVO conditional on a set of variables, and f(t) is a flexible function of time *t*.

When the dependent variables are all set to zero, the baseline hazard rate can be written as a function of time duration t,  $h_0(t) = \frac{e^{f(t)}}{1+e^{f(t)}}$ . f(t) allows the baseline hazard rate of receiving a CGVO to vary over time t. In effect, the logit model has the following form:

(2) 
$$\log\left(\frac{P_{it}}{1-P_{it}}\right) = \beta_0 + \beta_1 * X_{i,2006} + \beta_2 * Z_i + f(t) + \varepsilon_{it}$$

where  $P_{it}$  is the probability of receiving a CGVO for village i at time t,  $X_{i,2006}$  are the timeinvariant welfare measures in 2006 (a year before the CGVO program), and  $Z_i$  are the timeinvariant basic village characteristics. Time duration f(t) is approximated by a 4th order polynomial function of t.<sup>2</sup>

<sup>&</sup>lt;sup>1</sup> Traditional logit or probit models assume duration independence, i.e. the probability of being treated at any point in time is always the same. This is not a valid assumption here because the probability of receiving a CGVO increases over time. Without taking into account duration dependence, the standard errors estimated from a traditional logit or probit model would be wrong.

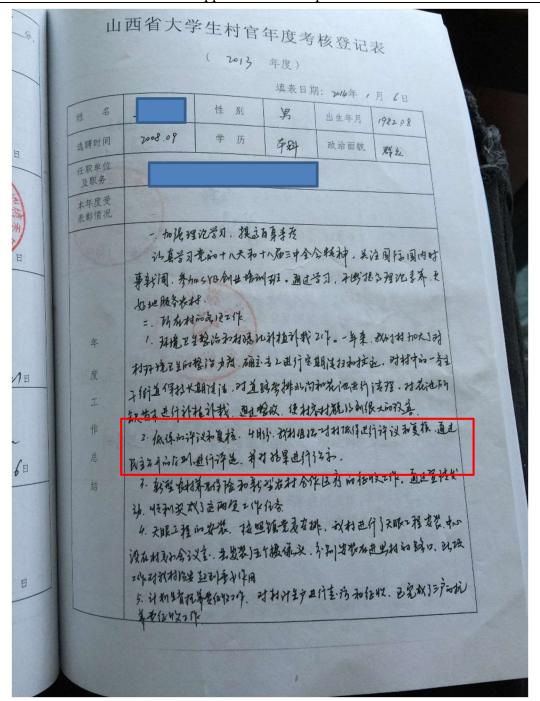
<sup>&</sup>lt;sup>2</sup> Approximating the time duration using a non-parametric method generates similar findings. The results are available upon request.

In columns (5) to (8) of Appendix Table 1, we include the same set of variables as in columns (1) to (4). The findings remain the same: none of these pre-determined village conditions have any effect, indicating that the assignment of CGVOs is likely to be exogenous to the village.<sup>3</sup>

In this longitudinal setting, we can also test the second hypothesis – whether CGVO assignment depends on village-level economic shocks – by including time-varying covariates in the regressions. Appendix Table 2 summarizes the results. The independent variables are changes in village population, income, poor housing, subsidized population, registered poor households, government size and quality of local government officials before the introduction of the CGVO program. None of these variables are statistically significant at a conventional level, indicating that economic shocks before the CGVO program did not affect CGVO assignments.

Whether the assignment decision is driven by time-varying shocks is critical to subsequent impact analysis. To identify causal effects, our main econometric model relies on variations in CGVO assignments across time and place in a difference-in-differences (DID) setting. The results in Appendix Table 2 confirm that CGVO assignments are not correlated with observed time-varying factors, suggesting that DID is likely to be a valid approach for estimating the impacts of the CGVO program.

<sup>&</sup>lt;sup>3</sup> The conclusions are the same if we use data from other years before 2006.

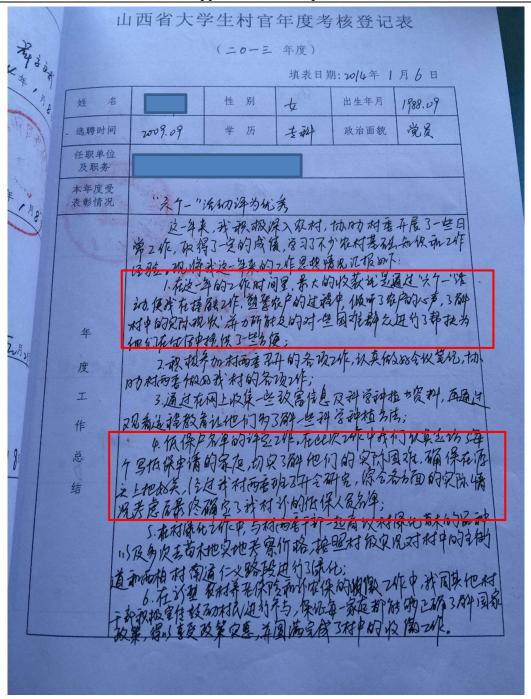


## Appendix B. CGVO Self-Evaluation Forms Appendix B1: Sample 1

## **Translation:**

Point 2 (Contribution to the Village): Select and double-check the Poverty Subsidy Applications. "In April (2013), I helped select and double-check the eligibility of the poverty-subsidy applicants. The beneficiaries were democratically determined by group voting, and the results were publicized to the entire village."

*Notes:* This form is used by Shanxi Province to evaluate the CGVO performance in 2013.



Appendix B2. Sample 2

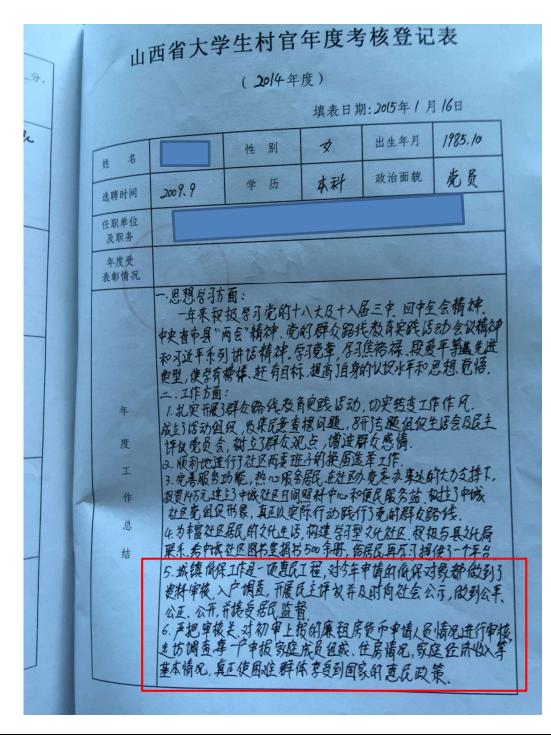
# **Translation:**

Point 1. "Over the past year, I have become more familiar with the conditions of the villagers and better understood their needs through deep conversations with them. I tried to offer some help to those who really have difficulties in life."

Point 4. "When deciding the beneficiaries of the poverty subsidy, I visited every applicant's home and collected detailed information on their living conditions. We held a village committee meeting and finalized the list of beneficiaries."

*Notes:* This form is used by Shanxi province to evaluate CGVO performance in 2013.

Appendix B3. Sample 3



## **Translation:**

Point 5. "For every poverty subsidy applicant, I strictly followed the procedures of screening application materials, conducting household surveys, organizing group evaluations, and publicizing results."

Point 6. "For all of the applicants for the government's subsidized housing program, I screened their materials, conducted household surveys, and especially focused on checking their current housing conditions, demographic compositions, and financial situations."

Notes: This form is used by Shanxi province to evaluate CGVO performance in 2014.

Appendix C. Village Condition Notebooks

#### 到织部 大学生村官工作资料汇编 大学生村官工作 月 日 星期 有了问题,我和孙子们的生活就 天气 访农户姓名 星期 能回到来 没有了满落。想让政府考虑我们 走访农户姓名: 联系方式 访农户基本 家的情况都能够申请低保。你回去 走访农户基本情况: 「家中王口人, 母亲妻 面有114 瘟症木 子和-双儿女。母亲年近,妻子多年闲 给问问。我给出的解释回答是我把你们 汤 蕃 离婚,把家中的有钱者引导走了。车 家的.情况修民政办反映,看是否符定低 方面没有 在一人 人因离媚患有神般病。一病起 保政策。 来就胡打人。王间平房 - []] 赤羽农 3. 孙子现在在哪上学,"学习怎么样? 走访内容: /、四公了家的具体清淡底样? 走访内容: 意解决。 秋于现在在咱村学校上学,学习还 前几年,在镁厂干活,被机器削 司以。 掉了一只耳朵。这几年在家务农,因为没 液。 亮百姓 北湖北,希望夫妻二人能够多考虑 钱.媳妇珍着爱闹离婚.把家中值钱 期外 名大学生 孩子,不能因为没钱,你们便闲 的都拿走了, 私人受了后刺激, 有些精 解子民语, 得要离婚,不能孩子们的感受,让 还说 神病,一直要挥母亲走,不让母亲住, 他们的小的公灵上得到创伤。 非弹 医尽自 把母亲房子的玻璃都敲的桥巴 相後只要你们去妻,齐心切办力.些 在的好 X兰。 没 送金度得美好。 2. 家中有什么 困难? 1:3 走访户签字: 你看, 因妻子 雨离媚, 儿子精神 P

# Appendix C1. Sample 1

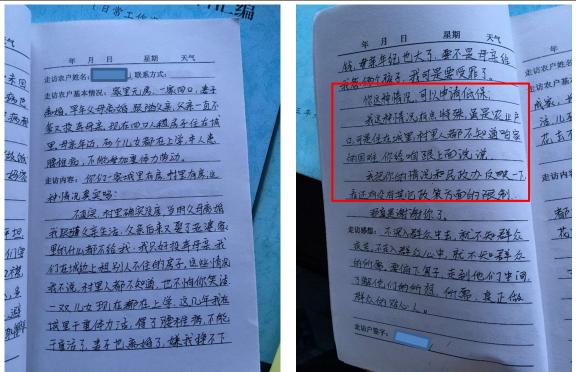
## **Translation:**

The villager said: "Please help relay the actual conditions of our family to the government and ask them whether we qualify for subsidies."

The CGVO responded: "I will inform the local Bureau of Civil Affairs about your conditions and see whether you qualify for the subsidy programs."

*Notes*: The village condition notebooks were used by CGVOs to record their daily work and document villagers' living conditions. These documents are archived by the Organization Department of the Central Committee of the Communist Party of China.

# Appendix C2. Sample 2



## **Translation:**

The CGVO advised the villager: "Given your conditions, you should consider applying for the poverty subsidy."

The villager responded: "My case is a special one. Although I am a rural resident, my house is in the suburban areas close to the city, so the villagers are not familiar with my real conditions and don't really understand my difficulties. Please help relay my information to the government."

*The CGVO responded: "I will talk to the local Bureau of Civil Affairs and see what they can do."* 

*Notes*: The village condition notebooks were used by CGVOs to record their daily work and villagers' living conditions. These documents are archived by the Organization Department of the Central Committee of the Communist Party of China.

	CGVO Assignment								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
Population	0.18	0.20	0.07	0.17	0.18	0.22	0.12	0.26	
	(0.23)	(0.31)	(0.36)	(0.37)	(0.20)	(0.28)	(0.28)	(0.30)	
Per capita Income	-0.05	-0.32	-0.40	-0.32	-0.03	-0.32	-0.43	-0.41	
	(0.28)	(0.43)	(0.45)	(0.50)	(0.27)	(0.43)	(0.45)	(0.47)	
Poor Housing		0.04	0.02	0.09		-0.01	0.00	0.04	
		(0.20)	(0.20)	(0.22)		(0.18)	(0.18)	(0.19)	
Subsidized Population		-0.21	-0.22	-0.15		-0.21	-0.21	-0.11	
		(0.40)	(0.40)	(0.42)		(0.34)	(0.35)	(0.37)	
Registered Poor HHs		-0.05	-0.05	-0.13		-0.11	-0.11	-0.21	
		(0.32)	(0.32)	(0.33)		(0.26)	(0.26)	(0.28)	
Government Size			0.07	0.05			0.05	0.04	
			(0.10)	(0.09)			(0.04)	(0.04)	
Government Quality			-0.00	-0.01			0.00	-0.00	
			(0.01)	(0.01)			(0.01)	(0.01)	
Terrain				0.38				0.29	
				(0.41)				(0.35)	
Pillar Industry				0.25				0.59	
				(0.79)				(0.69)	
Suburb				0.24				0.23	
				(0.50)				(0.39)	
Town Center				-0.30				-0.16	
				(0.43)				(0.34)	
Designated Poor Village				0.04				-0.08	
				(0.78)				(0.62)	
Precipitation				-59.17				-98.75	
				(112.64)				(108.80)	
Temperature				0.01				0.01	
				(0.02)				(0.02)	
Time Duration			-		4th Order Polynomial				
Psudo $R^2$	0.00	0.01	0.02	0.03	0.19	0.21	0.21	0.22	
Observations	233	143	143	143	2,421	1,479	1,476	1,476	

Appendix Table A1. Probability of CGVO Assignment: Pre-CGVO Levels

*Notes* : The probability of CGVO assignment is estimated using logit models. In columns (1) - (4), we estimate cross-sectional regressions in which the dependent variable is the eventual treatment status and the independent variables are village characteristics in 2006. Robust standard errors are reported in parentheses. In columns (5)-(8), we estimate the associations using a logit model with duration dependence with the panel data. We include a fourth order polynomial function to approximate the duration. Standard errors are clustered at the village level and reported in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

		C	CGVO A	ssignme	nt	
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta$ in Village Population	0.30	0.37	0.31	0.31	0.32	0.30
(by 1000)	(0.27)	(0.29)	(0.27)	(0.28)	(0.27)	(0.27)
$\Delta$ in per capita Income	0.19	0.13	0.37*	0.15	0.19	0.20
(by 1000 yuan)	(0.15)	(0.16)	(0.19)	(0.16)	(0.14)	(0.14)
$\Delta$ in the Share of Poor Housing		-0.93				
(by 100)		(0.59)				
$\Delta$ in Subsidy Rate			0.33			
(by 100)			(0.24)			
$\Delta$ in the Share of Registered Poor HHs				0.21		
(by 100)				(1.03)		
$\Delta$ in Government Size					-7.72	
(by 100)					(7.47)	
$\Delta$ in Government Quality						0.04
(by 100)						(0.92)
Time Duration		4tl	h Order 1	Polynom	ial	
Psudo R <sup>2</sup>	0.15	0.15	0.18	0.15	0.15	0.15
Observations	1,803	1,463	1,184	1,660	1,799	1,799

Appendix Table A2. Probability of CGVO Assignment: Pre-CGVO Shocks

*Notes:* The probability of CGVO assignment is estimated using logit models with duration dependence. We include a fourth order polynomial function to approximate the duration. The independent variables are changes in socioeconomic conditions before the CGVO program. Standard errors are clustered at the village level and reported in parentheses. \*\*\* p < 0.01, \*\* p < 0.05, \* p < 0.1.

	Subsidized Population (per 1000,					Poor Housing (per 100 households,				
		1	og)				log)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)		
CGVO	0.18**	0.18**			-0.07	-0.07				
	(0.07)	(0.07)			(0.05)	(0.05)				
	(0.12)	(0.12)			(0.05)	(0.05)				
	(0.10)	(0.10)			(0.07)	(0.07)				
L.CGVO			0.20***	0.20***			-0.15***	-0.15***		
			(0.07)	(0.07)			(0.05)	(0.06)		
			(0.12)	(0.12)			(0.08)	(0.10)		
			(0.11)	(0.11)			(0.10)	(0.10)		
Controls	Ν	Y	Ν	Y	N	Y	Ν	Y		
Village FE	Y	Y	Y	Y	Y	Y	Y	Y		
P-Y FE	Y	Y	Y	Y	Y	Y	Y	Y		
Obs.	2,102	2,102	2,102	2,102	2,417	2,417	2,417	2,417		
R <sup>2</sup>	0.67	0.67	0.67	0.67	0.78	0.78	0.78	0.78		

Appendix Table D1. Robustness Checks: CGVO and Subsidies

*Notes*: This table estimates the impacts of CGVOs on poverty subsidies and poor-quality housing using within province variation in CGVO assignment. We include village fixed effects and province-year fixed effects in all regressions. Below the estimated coefficients are standard errors clustered at the province-year, provincial and village level respectively. The asterisks indicate significance levels corresponding to standard errors clustered at the province-year level. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

	Register		ouseholds g)	People with Disabilities (per 1000, log)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CGVO	0.16***	0.16***			0.08	0.08		
	(0.05)	(0.05)			(0.07)	(0.07)		
	(0.05)	(0.05)			(0.08)	(0.08)		
	(0.08)	(0.08)			(0.07)	(0.08)		
L.CGVO			0.20***	0.20***			0.12*	0.13*
			(0.06)	(0.06)			(0.07)	(0.07)
			(0.07)	(0.07)			(0.10)	(0.10)
			(0.08)	(0.08)			(0.10)	(0.10)
Controls	Ν	Y	Ν	Y	N	Y	Ν	Y
Village FE	Y	Y	Y	Y	Y	Y	Y	Y
P-Y FE	Y	Y	Y	Y	Y	Y	Y	Y
Obs.	2,654	2,654	2,654	2,654	1,826	1,826	1,826	1,826
$\mathbb{R}^2$	0.69	0.69	0.69	0.69	0.75	0.75	0.75	0.75

Appendix Table D2. Robustness Checks: Registration Effect

*Notes*: This table estimates the impacts of CGVOs on registered poor households and people with disabilities using within province variation in CGVO assignment. We include village fixed effects and province-year fixed effects in all regressions. Below the estimated coefficients are standard errors clustered at the province-year, provincial and village level respectively. The asterisks indicate significance levels corresponding to standard errors clustered at the province-year p<0.05, \* p<0.1.

	Subsidized Population (per 1000, log)				Poor Housing (per 100 households, log)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
CGVO	0.20*	0.19			-0.07	-0.07		_	
	(0.11)	(0.11)			(0.05)	(0.05)			
	(0.10)	(0.10)			(0.07)	(0.07)			
	(0.11)	(0.11)			(0.07)	(0.07)			
L.CGVO			0.23**	0.22**			-0.13*	-0.13*	
			(0.10)	(0.10)			(0.07)	(0.07)	
			(0.11)	(0.11)			(0.08)	(0.08)	
			(0.11)	(0.11)			(0.08)	(0.08)	
Controls	Ν	Y	Ν	Y	N	Y	Ν	Y	
Village FE	Y	Y	Y	Y	Y	Y	Y	Y	
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	
Obs.	2,067	2,067	2,067	2,067	2,376	2,376	2,376	2,376	
$\mathbb{R}^2$	0.62	0.62	0.62	0.62	0.75	0.75	0.76	0.76	

Appendix Table E1. Robustness Checks: Dropping Villages with CGVOs before 2007

*Notes:* This table estimates the impacts of CGVOs on poverty subsidies and poor-quality housing. We exclude villages that received CGVOs before 2007 from the sample. We probe the robustness of estimate accuracy by clustering the standard errors at three different levels: provincial, village, and village and province-year level (multi-way clustering suggested by Cameron, Gelbach, and Miller (2011)). These standard errors are respectively reported in the parentheses below the estimated coefficients. Our preferred specification clusters standard errors at the provincial level. As we only have 19 provinces, we address the small sample bias in the clustered standard errors using wild bootstrapping, a method recommended by Cameron, Gelbach and Miller (2008). The significance levels indicated by asterisks are based on wild bootstrapped p-values, which are similar to the simple significance levels using standard errors clustered at the provincial level. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

	Registered Poor Households (per 100, log)				People with Disabilities (per 1000, log)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
CGVO	0.09	0.09			0.09	0.09			
	(0.06)	(0.06)			(0.07)	(0.07)			
	(0.07)	(0.07)			(0.07)	(0.07)			
	(0.08)	(0.08)			(0.08)	(0.08)			
L.CGVO			0.15**	0.15**			0.16*	0.16*	
			(0.07)	(0.07)			(0.09)	(0.09)	
			(0.08)	(0.08)			(0.09)	(0.09)	
			(0.09)	(0.09)			(0.10)	(0.09)	
Controls	Ν	Y	Ν	Y	N	Y	Ν	Y	
Village FE	Y	Y	Y	Y	Y	Y	Y	Y	
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	
Obs.	2,608	2,608	2,608	2,608	1,792	1,792	1,792	1,792	
$\mathbb{R}^2$	0.65	0.65	0.65	0.65	0.73	0.73	0.73	0.73	

Appendix Table E2. Robustness Checks: Dropping Villages with CGVOs before 2007

*Notes:* This table estimates the impacts of CGVOs on registered poor households and people with disabilities. We exclude villages that received CGVOs before 2007 from the sample. We probe the robustness of estimate accuracy by clustering the standard errors at three different levels: province, village, and village and province-year level (multi-way clustering suggested by Cameron, Gelbach, and Miller (2011)). These standard errors are respectively reported in the parentheses below the estimated coefficients. Our preferred specification clusters standard errors at the provincial level. As we only have 19 provinces, we address the small sample bias in the clustered standard errors using wild bootstrapping, a method recommended by Cameron, Gelbach and Miller (2008). The significance levels indicated by asterisks are based on wild bootstrapped p-values, which are similar to the simple significance levels using standard errors clustered at the provincial level. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

	Subsidized Population (per 1000, log)				Poor Housing (per 100 households, log)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
CGVO	0.20*	0.20*			-0.09*	-0.09**			
	(0.11)	(0.11)			(0.04)	(0.04)			
	(0.10)	(0.10)			(0.07)	(0.07)			
	(0.11)	(0.11)			(0.07)	(0.07)			
L.CGVO			0.20*	0.20*			-0.14**	-0.14**	
			(0.10)	(0.10)			(0.06)	(0.06)	
			(0.10)	(0.10)			(0.08)	(0.08)	
			(0.10)	(0.11)			(0.08)	(0.08)	
Controls	Ν	Y	Ν	Y	N	Y	Ν	Y	
Village FE	Y	Y	Y	Y	Y	Y	Y	Y	
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	
Obs.	2,102	2,102	2,102	2,102	2,417	2,417	2,417	2,417	
R <sup>2</sup>	0.61	0.61	0.61	0.61	0.76	0.76	0.76	0.76	

Appendix Table F1. Robustness Checks: Using Alternative CGVO Dummy

*Notes:* This table estimates the impacts of CGVOs on poverty subsidies and poor-quality housing using an alternative CGVO treatment dummy. In these regressions, a village is considered treated starting from the first year it received a CGVO, and until the end of our study period in 2011, regardless of whether a CGVO left a village during the period. We probe the robustness of estimate accuracy by clustering the standard errors at three different levels: provincial, village, and village and province-year level (multi-way clustering suggested by Cameron, Gelbach, and Miller (2011)). These standard errors are respectively reported in the parentheses below the estimated coefficients. Our preferred specification clusters standard errors at the provincial level. As we only have 19 provinces, we address the small sample bias in the clustered standard errors using wild bootstrapping, a method recommended by Cameron, Gelbach and Miller (2008). The significance levels indicated by asterisks are based on wild bootstrapped p-values, which are similar to the simple significance levels using standard errors clustered at the provincial level. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

	Registered Poor Households (per 100, log)				People with Disabilities (per 1000, log)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
CGVO	0.10*	0.10*			0.12	0.11			
	(0.06)	(0.06)			(0.07)	(0.07)			
	(0.07)	(0.07)			(0.07)	(0.07)			
	(0.08)	(0.08)			(0.08)	(0.08)			
L.CGVO			0.14**	0.14**			0.16*	0.16*	
			(0.06)	(0.06)			(0.09)	(0.09)	
			(0.08)	(0.08)			(0.09)	(0.08)	
			(0.09)	(0.09)			(0.09)	(0.09)	
Controls	Ν	Y	Ν	Y	N	Y	Ν	Y	
Village FE	Y	Y	Y	Y	Y	Y	Y	Y	
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	
Obs.	2,654	2,654	2,654	2,654	1,826	1,826	1,826	1,826	
<b>R</b> <sup>2</sup>	0.65	0.65	0.65	0.65	0.73	0.73	0.73	0.73	

Appendix Table F2. Robustness Checks: Using Alternative CGVO Dummy

*Notes:* This table estimates the impacts of CGVOs on registered poor households and people with disabilities using an alternative CGVO treatment dummy. In these regressions, a village is considered treated starting from the first year it received a CGVO and until the end of our study period in 2011, regardless of whether a CGVO left a village during the period. We probe the robustness of estimate accuracy by clustering the standard errors at three different levels: provincial, village, and village and province-year level (multi-way clustering suggested by Cameron, Gelbach, and Miller (2011)). These standard errors are respectively reported in the parentheses below the estimated coefficients. Our preferred specification clusters standard errors at the provincial level. As we only have 19 provinces, we address the small sample bias in the clustered standard errors using wild bootstrapping, a method recommended by Cameron, Gelbach and Miller (2008). The significance levels indicated by asterisks are based on wild bootstrapped p-values, which are similar to the simple significance levels using standard errors clustered at the provincial level. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.