Learning to love big brother: Chinese attitudes toward online privacy after the pandemic

Andrew MacDonald

2024-09-04

test

1 Introduction

2 Literature review

2.1 Hypotheses

3 Data and summary statistics

The data for this project was collected via a commercial survey firm in two waves, February of 2021 and March of 2023. In both the first and second waves, Wuhan was oversampled, with residents of the city set to be 10% of respondents. The 2021 survey had an n=1500 and the second had an n=2000. Questions on the two surveys were identical other than a minor change to a question that referenced a specific date. The timing of the two surveys came at two very different points in time of China's Covid-19 experience. The first survey was conducted approximately seven months after the last round of restrictions were lifted on the city of Wuhan. China, at the time, was essentially closed to foreign travel but otherwise had little in the way of day to day public health restrictions. Nationwide, daily Covid cases hovered around the single digits (*BBC News*, 2021). China was at a very different point in its journey in March of 2023. The year of 2022 saw widespread, intrusive digital monitoring introduced. Many major cities, such as Shanghai, Xi'an, and Shenzhen, underwent long and painful city-wide lockdown procedures. At the end of 2023, under the weight of a spiraling number of cases and widespread protests (termed the White Paper Revolution), China finally abandoned its zero Covid policy (Mao, 2022). The two waves of these surveys aim to compare attitudes before and after this widespread and highly visible change in digital monitoring strategies.

The demographics of the 2021 and 2023 surveys are presented in Table 1.

Table 1: Select key demographic variables

		Mean	Std. Dev.
Age		33.2	11.6
		N	Pct.
Location	Countryside/village	477	13.6
	Small city	1059	30.2
	Mid-sized city	840	24.0
	Big city	1131	32.2
Education	No formal education	22	0.6
	Primary	134	3.8
	Middle school	384	10.9
	High school	843	24.0
	University	1914	54.6
	Advanced studies/Graduate school	210	6.0
Gender	Female	1711	48.8
	Male	1796	51.2
Marriage status	Single	1101	31.4
	In a relationship	569	16.2
	Married	1744	49.7
	Divorced	93	2.7
Party member status	Yes	483	13.8
	No	3024	86.2
Communist Youth League status	Yes	1116	31.8
	No	2391	68.2
Income	0-2,999	275	7.8
	3,000-5,999	822	23.4
	6,000-9,999	899	25.6
	10,000-19,999	962	27.4
	20,000-49,999	385	11.0
	50,000-99,999	94	2.7
	More than 100,000	70	2.0
Year	2021	1500	42.8
	2023	2007	57.2

As is typical of online surveys in China, the sample respondents skew somewhat younger and more educated. Comparing the two waves, there are some modest demographic differences (notably education and marriage) differences between the two samples. As will be shown in Section 4, these minor differences do not appear to change any of the substantive results. Focusing on the 2023 survey, the modal respondent is someone from a small city, male, married,

Table 2: Questions asking about attitudes toward government monitoring

GM1	There are good reasons for the central government to monitor the activity of users online
GM2	There are good reasons for the local government to monitor the activity of users online
TRACK1	How comfortable are you with the central government knowing personal details about your activity online?
TRACK2	How comfortable are you with the local government knowing personal details about your activity online?

working in a white collar job at a small enterprise, who earns about 10,000 RMB a month and has an urban *hukou*.

To simplify the analysis that follows, the variables are recoded such that income is divided into three categories (low, middle, and high) and education is divided into two categories, those with college education and those without. The regressions in the following sections were tested with alternate specifications of these categorical variables (code for these regressions available at the author's webite) and the key results were unaffected.

The key response variable for the following analysis is an index variable created by combining the results of the four questions in Table 2, rescaled to be between zero and one. It is true that previous research has found that Chinese respondents place lower levels of trust in local governments as compared to central governments (CITE). However, the correlation between `GM1` and `GM2` is 0.84 and the correlation between `TRACK1` and `TRACK2` is 0.87. Additionally, as shown in the online appendices, each of the major regression results in Section 4 do not demonstrate major changes if the same variables are regressed on each of the survey question items individually. Indicating their close relationship, the variables taken together have a Cronbach's α of 0.81. Intuitively, this high level of relatedness makes sense as while respondents have some background attitudes about the difference between central and local governments, they may not easily be able to identify at which level government tracking occurs.

Finally, there has been some debate as to the extent of preference falsification on online surveys in China. To test for preference falsification, the surveys also contained several list experiment questions. List experiments have been used in a number of surveys (CITE) to allow residents a confidential method to express their true attitudes about topics such as racial views, sexual assault experience, and political views. While (CITE) points out that list experiments should not be seen as silver bullet to the problem of preference falsification, the results of the list experiments (also available in the online appendices) roughly match the answers to the component questions that comprise the response variable index.

Overall, the survey data should provide a robust test to arbitrate between the hypotheses posed in Section 2.

4 Analysis

4.1 Predicting citizen attitudes about government monitoring

The first set of models considers the question of which demographic variables predict variation in attitudes toward being tracked by the government. (stuff related to lit review)

The key predictor variables included in Table 3 are the tech savvy index (TSI) and the knowledge index (KI). The items used to construct each are listed in Section 7. The Cronbach's α for each set of variables is 0.8 and 0.55 respectively, indicating that the questions are suitable for use in an index.

The coefficients generally indicate effects in the direction expected. Older respondents are more accepting of government tracking, while being male and not being a member of the party predict lower acceptance of tracking. Curiously, living in a big city has a positive relationship with tracking acceptance, as does the year 2023. Being tech savvy is positively related to acceptance of tracking, while having knowledge of privacy is negatively associated with the response variable.

However, with respect to the magnitude of the coefficients, the response variable is scaled between zero and one with a standard deviation of 0.2. Given this scaling, the coefficients of the categorical variables all have rather small effect sizes - being in year 2023 instead of year 2021 produces a shift in the response variable of about a third of a standard deviation. For the other categorical variables, while some reach significance, they have even smaller effect sizes. The tech savvy index has a standard deviation of 0.21. The coefficient of TSI indicates the impact of a one unit change in the index (going from its minimum to its maximum) on the response variable. However, a more typical shift in TSI produces an effect only one fifth as large, or about a fifth to a tenth of a standard deviation change in the response variable. Similarly, for KI, a typical shift in the predictor variable leads to a nearly negligible change in the response variable.

As can be inferred from the results in this table, the model fit is relatively poor. The poor model fit can also be seen in the very low R^2 values and in the extremely poor model residuals (available in the online appendices). Taken together, these results indicate that the available demographic factors do a poor job explaining variation in attitudes towards government tracking.

4.2 Does government trust affect attitudes?

Another plausible relationship is that generalized trust in government is a strong predictor of attitudes toward government monitoring. Similar to the key variables in the preceding section, the measure of both local government and central government performance are combined into an index (GPI). As noted in Section 3, while there has been an observed gap in measurement

Table 3: Demographic predictors of attitude toward government privacy

	(1)	(2)	(3)	(4)	(5)	(6)
(Intercept)	0.573***	0.548***	0.488***	0.506***	0.504***	0.576***
	(0.019)	(0.019)	(0.021)	(0.023)	(0.022)	(0.022)
Age	0.001***	0.001**	0.002***	0.001***	0.002***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
College education	0.010	0.014+	0.006	−0.026	0.006	0.014+
	(0.008)	(0.008)	(0.008)	(0.017)	(0.008)	(0.008)
Middle income	0.000	0.002	−0.003	−0.004	−0.003	0.001
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
High income	−0.015	−0.012	−0.024	−0.025	−0.024	−0.014
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Male	−0.004	−0.003	−0.009	−0.010	−0.040*	−0.004
	(0.007)	(0.007)	(0.007)	(0.007)	(0.016)	(0.007)
Not a party member	−0.033**	−0.034***	−0.030**	−0.029**	−0.030**	−0.032**
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Location: small city	0.000	0.002	−0.003	−0.001	−0.003	0.002
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Location: mid city	0.007	0.009	0.000	0.001	0.000	0.008
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Location: big city	0.013	0.014	0.001	0.002	0.001	0.013
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Year 2023		0.037***	0.039***	0.038***	0.038***	0.036***
		(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
TSI			0.118***	0.010	0.017	
			(0.017)	(0.055)	(0.052)	
TSI x education				0.069*		
				(0.034)		
TSI x sex					0.065*	
					(0.032)	
KI						−0.058**
						(0.022)
Num.Obs.	3507	3507	3507	3507	3507	3507
R2	0.008	0.016	0.029	0.030	0.030	0.018
F	3.271	5.846	9.539	9.105	9.107	5.928

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Reference values: no college education, low income, female, party member, countryside

Standard deviation of the response variable: 0.2

of the two concepts in previous literature (and in these surveys), the responses to the two questions are nevertheless highly correlated ($\rho = 0.73$). To the extent that they measure disjoint opinions, models 1b and 1c take central government performance alone and local government performance alone, as the response variables.

The coefficients on in Table 4 indicate that government performance is a much stronger predictor of attitudes towards government tracking than the index demographic variables. In model 1, a one standard deviation increase in the government performance index (0.2) predicts about a one third of a standard deviation change in attitudes towards government tracking, a relatively significant effect for an attitudinal survey. Additionally, the two interaction terms are also both significant. The effect of these interaction terms can be viewed in Figure 1. Education lessens the impact of government performance on acceptance of government tracking while year 2023 increases the impact. Finally, the model fit diagnostics have improved, indicating a better model fit.

As expected, government performance, controlling for demographic factors, is a relatively strong predictor of acceptance of government performance. According to the model, a positive view of government performance is associated with an increased acceptance of government monitoring. However, the relationship is likely more complex than this story, particularly given the tumultuous events of the 2022 Covid-19 lockdowns in China. To further explore how these factors interact with each other, the next section develops a mediation model to better understand the causal factors at play.

4.3 Changing attitudes since the pandemic

The following directed acyclic graph (DAG) indicates the hypothesized causal process that generates the observed outcome variable, tracking acceptance. In Figure 2, `TA` represents tracking acceptance, `GP` represents government performance approval, `DEMO` represents demographic characteristics, and `COVID` represents respondents' Covid-19 experience.

To operationalize this model, the first step is to create a latent demographics variable with the demographic factors previously used in regressions in the earlier sections loading onto this variable. `COVID` is operationalized by the `year` variable. Admittedly, this is not a precise mapping. The `year` variable actually measures all changes between survey waves not accounted for by other variables. With respect to government tracking acceptance attitudes, however, this assumption can be justified by the fact that the Covid-19 experience was both a daily and often traumatic one for the Chinese public; it was a time period that involved constant and invasive technological monitoring. If the model does indicate that the `year` variable predicts significant change in the response variable over the two year difference between survey waves, it would be hard to imagine any other plausible cause. Nevertheless, it is important to keep in mind that it is only a proxy measurement. These variables (`year` for `COVID`, `GP`, `TA`, and `DEMO`) are entered into a structural equation model with paths constrained to that described by the DAG in Figure 2 and then the parameters are estimated using the `lavaan` library in `R`.
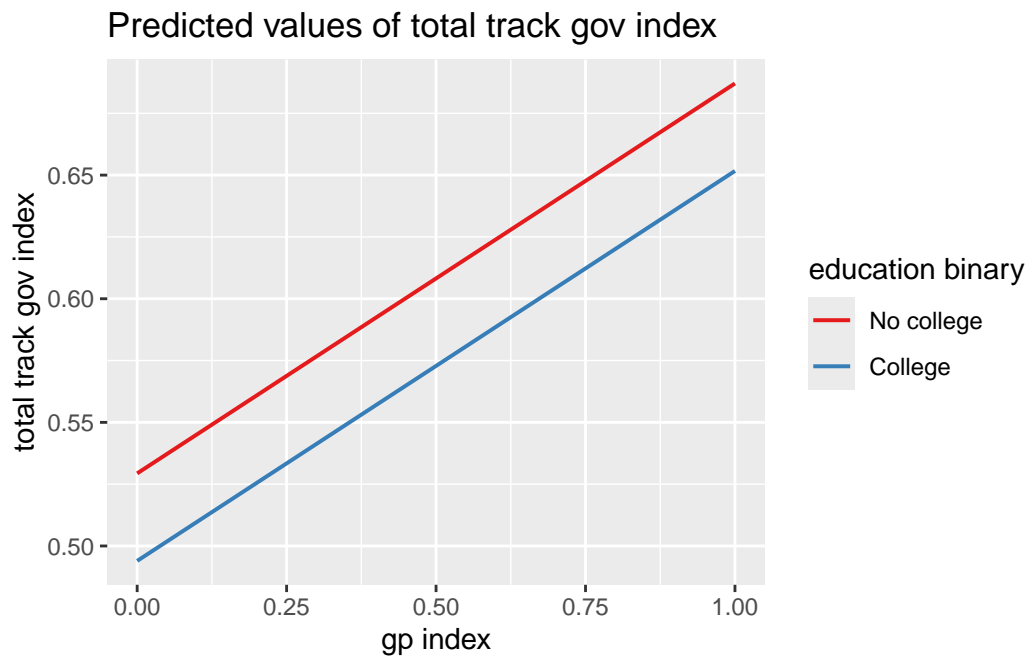
Table 4

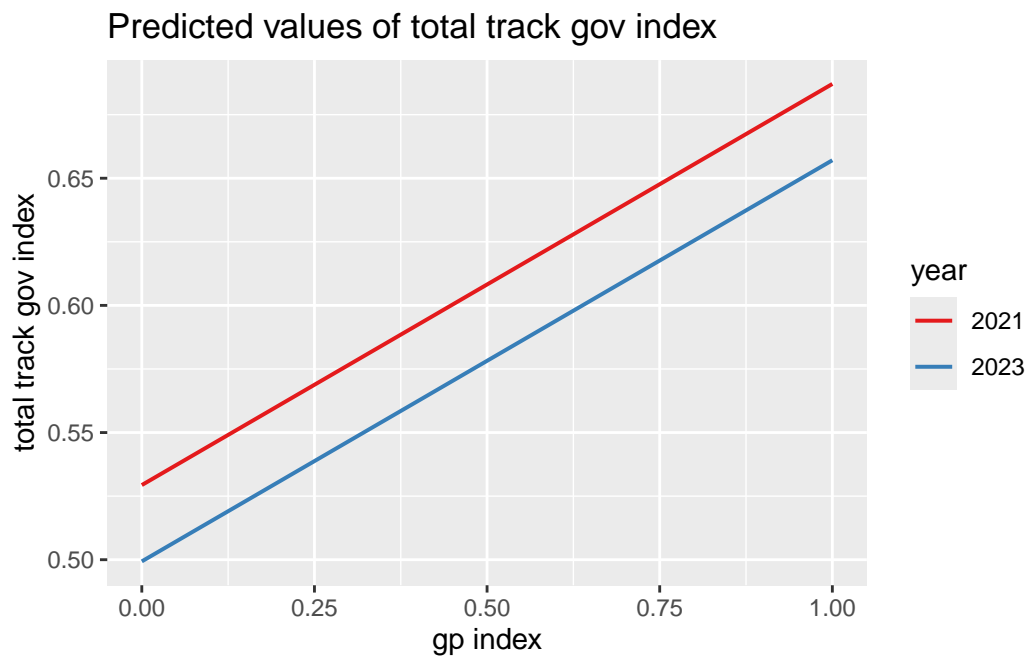|	(1)	(1a)	(1b)	(2)	(3)	(4)
(Intercept)	0.214***	0.212***	0.299***	0.236***	0.266***	0.297***
	(0.021)	(0.021)	(0.020)	(0.024)	(0.027)	(0.031)
Age	0.001**	0.001**	0.001***	0.001***	0.001**	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
College education	0.012+	0.012+	0.013+	−0.026	0.012+	−0.035
	(0.007)	(0.007)	(0.007)	(0.024)	(0.007)	(0.024)
Middle income	−0.003	−0.001	−0.003	−0.003	−0.003	−0.003
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
High income	−0.013	−0.010	−0.015	−0.013	−0.013	−0.013
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Male	−0.005	−0.006	−0.003	−0.005	−0.005	−0.005
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Not a party member	−0.028**	−0.026**	−0.031***	−0.028**	−0.029**	−0.029**
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Location: small city	−0.006	−0.003	−0.007	−0.006	−0.006	−0.006
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Location: mid city	−0.001	0.008	−0.006	−0.001	−0.001	−0.001
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Location: big city	0.011	0.022*	0.002	0.010	0.010	0.010
	(0.010)	(0.010)	(0.011)	(0.010)	(0.010)	(0.010)
Year 2023	0.052***	0.053***	0.047***	0.052***	−0.024	−0.030
	(0.006)	(0.006)	(0.006)	(0.006)	(0.025)	(0.025)
GPI	0.431***			0.352***	0.269***	0.158*
	(0.015)			(0.050)	(0.054)	(0.077)
CG performance		0.406***				
		(0.014)				
LG performance			0.344***			
			(0.014)			
GPI x education				0.050+		0.062*
				(0.030)		(0.030)
GPI x year					0.098**	0.105***
					(0.031)	(0.032)
Num.Obs.	3507	3507	3507	3507	3507	3507
R2	0.207	0.200	0.165	0.207	0.209	0.210
F	82.827	79.326	62.844	76.189	76.920	71.388

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Reference values: no college education, low income, female, party member, countryside

Standard deviation of the response variable: 0.2

Predicted values of total track gov index

(a) GPI x education

Predicted values of total track gov index

(b) GPI x year

Figure 1: Marginal effect plots of interaction terms

Table 5: Mediation model results

	(1)
DEMO to GP	−0.008
	(0.005)
GP to TA	0.430***
	(0.015)
DEMO to TA	−0.010*
	(0.005)
COVID to TA	0.053***
	(0.006)
COVID to GP	−0.037***
	(0.007)
DEMO to GP to TA	−0.003
	(0.002)
COVID to GP to TA	−0.002***
	(0.000)
Num.Obs.	3507
AIC	52 129.6
BIC	52 246.7

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

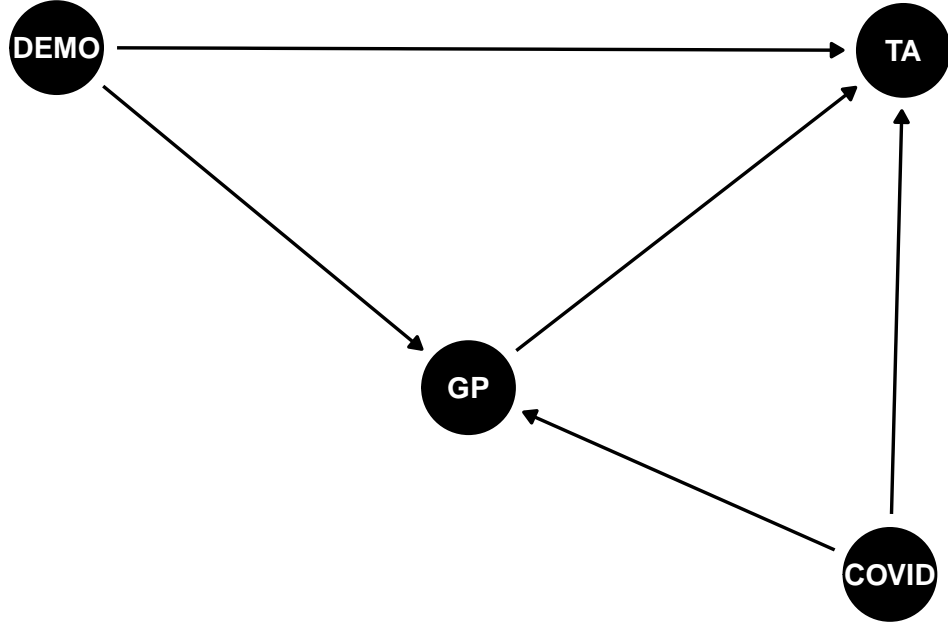Demographic factor variable loadings omitted

Figure 2: Causal process

The results from this analysis in Table 5 confirm some of the previous findings and also reveal some interesting new features of the data. As before, government performance plays an important role in determining tracking acceptance. Similarly, demographic questions do help predict tracking acceptance, but only by a little bit. Demographics do not help predict views on government performance, and therefore, it is not surprising that demographics does not have an indirect effect on tracking acceptance through government performance either. However, the Covid-19 experience, as operationalized by the year variable, does 1) positively increase tracking acceptance (direct path) 2) negatively decreases government performance evaluations (direct path) and 3) negatively decreases tracking acceptance through government performance evaluations (indirect path). The size of the coefficients indicates that the sum of the effects is still positive on tracking acceptance.

4.4 Comparing attitudes to private monitoring

Finally, it is interesting to compare the determinants of government tracking attitudes with those that determine attitudes toward private tracking of personal information. T compare these two, Table 6 includes models that use the previous acceptance of government tracking index (`Public TA`) as the variable alongside a similarly constructed index variable that measures acceptance of private monitoring (`Private TA`).

Table 6: Comparison of public vs. private tracking acceptance

	Public TA			Private TA		
	(1a)	(1b)	(1c)	(2a)	(2b)	(2c)
(Intercept)	0.488***	0.576***	0.214***	0.270***	0.432***	0.344***
	(0.021)	(0.022)	(0.021)	(0.025)	(0.026)	(0.027)
Age	0.002***	0.001***	0.001**	0.000	−0.001**	−0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
College education	0.006	0.014+	0.012+	−0.027**	−0.010	−0.010
	(0.008)	(0.008)	(0.007)	(0.009)	(0.009)	(0.009)
Middle income	−0.003	0.001	−0.003	−0.059***	−0.049***	−0.049***
	(0.008)	(0.008)	(0.007)	(0.009)	(0.009)	(0.009)
High income	−0.024	−0.014	−0.013	−0.027	−0.004	−0.002
	(0.017)	(0.017)	(0.015)	(0.019)	(0.020)	(0.020)
Male	−0.009	−0.004	−0.005	−0.006	0.006	0.007
	(0.007)	(0.007)	(0.006)	(0.008)	(0.008)	(0.008)
Not a party member	−0.030**	−0.032**	−0.028**	−0.010	−0.016	−0.017
	(0.010)	(0.010)	(0.009)	(0.012)	(0.012)	(0.012)
Location: small city	−0.003	0.002	−0.006	−0.043**	−0.032*	−0.033*
	(0.011)	(0.011)	(0.010)	(0.013)	(0.013)	(0.013)
Location: mid city	0.000	0.008	−0.001	−0.056***	−0.038**	−0.039**
	(0.012)	(0.012)	(0.011)	(0.014)	(0.014)	(0.014)
Location: big city	0.001	0.013	0.011	−0.051***	−0.024+	−0.024+
	(0.012)	(0.012)	(0.010)	(0.014)	(0.014)	(0.014)
Year 2023	0.039***	0.036***	0.052***	0.006	0.001	0.005
	(0.007)	(0.007)	(0.006)	(0.008)	(0.008)	(0.008)
TSI	0.118***			0.252***		
	(0.017)			(0.021)		
KI		−0.058**			−0.069**	
		(0.022)			(0.027)	
GPI			0.431***			0.071***
			(0.015)			(0.020)
Num.Obs.	3507	3507	3507	3507	3507	3507
R2	0.029	0.018	0.207	0.057	0.018	0.020
F	9.539	5.928	82.827	19.081	5.865	6.450

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Reference values: no college education, low income, female, party member, countryside

Standard deviation of Public TA is: 0.2

Standard deviation of Private TA is: 0.24

The results here suggest that the determinants of attitudes towards private company tracking are somewhat different than those that determine attitudes towards public tracking. The `year` variable is not significant for the private tracking models, while the tech savvy coefficient has roughly doubled. Unsurprisingly, government performance is also not related to private tracking acceptance. Finally, the intercept is generally lower for private tracking acceptance, indicating that respondents are less willing to accept private tracking, all things being equal.

5 Conclusion

6 References

BBC News (2021) Wuhan lockdown: A year of china's fight against the covid pandemic. Epub ahead of print 22 January 2021.

Mao (2022) China abandons key parts of zero-covid strategy after protests. *BBC News.* Epub ahead of print 7 December 2022.

Table 7: Tech savvy index questions

Q1	How would you rate your general ability to use a computer?
Q2	How would you rate your skill at fixing a computer?
Q3	How would you rate your ability to program a computer?

Table 8: Knowledge index questions

Q1	I am very concerned about my privacy online
Q2	I spend a lot of time reading about technology related privacy issues
Q3	In the last year, I have had discussions with my friends about online privacy issues
Q4	I feel like I know exactly how much privacy I have online
Q5	Have you heard of the social credit system (official terminology)?

7 Appendix

7.1 Index variable definitions

Table 9: Government performance index questions

Q1	Overall, I'm happy with the performance of the central government
Q2	Overall, I'm happy with the performance of my local government