The University of Melbourne

An Econometrics Analysis into the Causality Relationship Between Cadre Status and Household Income in China using 2SLS Techniques

Boxiang Fu

998092

CHIN20021

Analysis of Contemporary Chinese Society

Word Count: 3285

5th November 2021

Abstract

The question of the relationship between an individual's political capital and its wealth accumulation abilities is mainly established though a statistical correlative approach in the current literature. Contemporary research on its causality relationship has been minimal. Through a two-stage-least-squares (2SLS) framework, this paper explores this question through a causal approach. Using an individual's first cadre status and their offspring's cadre status as instrumental variables, this paper finds evidence that in politically-established geographical regions, there is a statistically significant causal effect of an individual's cadre status on the household's income and wealth generating abilities. Such information would prove helpful for policy makers, academics, and households for understanding issues regarding the efficiency of resource allocation.

Keywords

Political capital, wealth accumulation, 2SLS, causality analysis, household income, cadre status

1. Introduction

Contemporary literature studying the relationship between household income and cadre status in China mainly concerns itself with statistical correlations. Minimal attempts have been made for a causal interpretation of the matter. Such causal relationships could prove helpful for policy makers and academics in analyzing the impulse-response effects of policy changes and structural changes in society. It would also be valuable information to households regarding the optimality of resource allocation. In a previous literature review, the author found that in aggregate, the statistical correlation is ambiguous. However, using available field studies, it was found that in "politically-established" regions, the returns to household income is positively correlated with cadre status at all conventional significance levels (Xie et al., 2009; Xie and Jin, 2015). Through the use of the two-stage-least-squares (2SLS) approach, this paper seeks to explore this causality relationship on a subset of the general population. In particular, the author suspects that the statistical correlation found is causal at conventional significant levels.

A causality analysis into this relationship is novel in the literature. As such, no existing theories currently exist. Therefore, Section 2 provides a theory behind such causality relationships. It also provides some background regarding the status of contemporary literature regarding this issue. Section 3 then provides empirical evidence on the theory. This section is further subdivided into a methodology section, a data section, a results section, and an a posteriori validity check of the results. Section 4 discusses potential limitations and problems of the 2SLS approach. Section 5 provides conclusions and presents areas for future research.

2. Background

In a previous literature review, the question of whether families with considerable political capital held a comparative advantage in wealth accumulation was analyzed. Using an individual's cadre status as a proxy for political capital, the author found that in "politically-established" cities, the cadre status of an individual increases the wealth generating capabilities of the household at all conventional significance levels (Xie et al., 2009; Xie and Jin, 2015). The author defined "politically-established" cities as "a city with a persistent and rigid political structure under the CPC" (Fu, 2021). As such, there seems to exist a positive statistical correlation between wealth generation and cadre status in politically-established cities.

However, contemporary literature fails to explain the causality relationship between them. An important distinction exists between a causality interpretation and a statistical interpretation. While a statistical relationship between household income and cadre status exists, it does not necessarily mean that the cadre status of an individual causes increases in household income. Econometrics and statistical theory tells us that if problems such as omitted variables bias or simultaneous causality exists, then simple ordinary least squares (OLS) regressions fails to have a causal interpretation (Stock and Watson, 2015, pp. 365–375). Such problems almost certainly exist in current OLS-styled analyses in the literature. It would be quite easily postulated that variables that are correlated with both cadre status and household income exists but are omitted from the OLS regression (i.e. ability bias, charisma, etc.). Similarly, some degree of reverse causality is quite possible. As households increase their wealth, they may want to increase their political capital. Thus, household income may very well cause the cadre status of the individual to some degree. Such hypothetical stories tell us that a simple OLS approach in analyzing the causality relationship between household income and cadre status is insufficient. As such, a theory backed by methods to extract the causal effects is needed.[1]

A 2SLS approach for causal studies is well established in the literature. The theory and methodology behind the 2SLS regression is discussed in Section 3. As for theory regarding the causal relationship between household income and cadre status, the author hypothesizes that cadre

[1] An interesting bit of trivia is that the 2021 Nobel Memorial Prize in Economics was awarded to Joshua Angrist and Guido Imbens for their study of causal effects.

status causes an increase in household income. This stems from the author's postulation that the cadre status of an individual is a well-representation of the household's political capital. Conventional economic theory tells us that in politically-established regions, this capital would enter as a factor of production, thus generating economic rents to the holder. Such theory stands as the basis behind empirical studies done by Nee and Opper (2010) which only offer a statistical interpretation. The author sees no reason why this theory fails when a causal interpretation is used; as such, it forms our theoretical underpinning in our empirical analysis on causality.

Note here that this paper focuses primarily on politically-established cities due to the limitations of currently available field studies. Despite this, the methodology presented here can be extended to cover the general population should future field studies be made available and appropriate exogenous variables be found.

3. Empirical Findings

3.1 Methodology

To study the causal relationship between cadre status and household income, we use the conventional 2SLS approach common in the literature. As alluded to earlier, problems such as omitted variables bias and simultaneous causality inherently causes OLS parameters to deviate from their causal parameters (Stock and Watson, 2015, pp. 365–375). Heuristically, this would mean that the magnitude and/or directional effects of an individual's cadre status on their household income cannot be obtained from a simple OLS regression of cadre status on household income. This is because the omission of correlated variables such as ability, charisma, etc. in a regression model would bias and render the results from the OLS regression inaccurate. Similarly, the presence of reverse causality may indicate that household income causes cadre status. This again renders the OLS regression to be inaccurate as we are essentially regressing two separate regressions (household income on cadre status and cadre status on household income) using one regression.

Such problems can be mitigated using a 2SLS approach. The approach tries to find a variable (commonly called an instrumental variable) that is only correlated with our parameter of interest (cadre status) and not with any other variable. Then, in the first-stage of the two-stage-least-squares approach, we try to separate the "true" value of the cadre status from any non-causal effects caused by the problems mentioned earlier. We then discard the non-causal effects and only use the "true" values to run a second regression to capture the causality relationship of the "true" values and household income. In this way, effects from cadre status can be separated from other sorts of capital and a causal relationship of cadre status on household income can be established.

We can mathematically show the heuristic description given above by considering a simple data generating process

$$HouseholdIncome_i = \beta_0 + \beta_1 CadreStatus_i + U_i$$

Where β_0 and β_1 are unknown population parameters and U_i is a random error term.

If we simply take an OLS regression with $CadreStatus_i$ as the conditional expectation, we get

$$E(HouseholdIncome_i|CadreStatus_i) = E(\beta_0 + \beta_1 CadreStatus_i + U_i|CadreStatus_i)$$

$$= \beta_0 + \beta_1 CadreStatus_i + E(U_i|CadreStatus_i)$$

From this, if the error term (of potentially an amalgamation of omitted variables) is correlated with the cadre status of the individual, then the term $E(U_i|CadreStatus_i)$ varies with cadre status. Thus, the standard assumption for causal interpretation of OLS results is violated.

The 2SLS approach solves this problem by first regressing the endogenous (cadre status) variable against a set of instrumental variables(s) before running a second regression against the results from the first regression (Stock and Watson, 2015, pp. 470–488). An instrumental variable

(IV) is a random variable that satisfies both exogeneity and relevance. I.e., it must be correlated only with cadre status, and not with any other variable. Its mathematical definition for the IV Z_i satisfies both

$$E(U_i|Z_i) = 0$$

$$\widehat{CadreStatus}_i = E(CadreStatus_i|Z_i) = \pi_0 + \pi_1 Z_i, \qquad \pi_1 \neq 0$$

Regressing the data generating process on Z_i, we get

$$E(HouseholdIncome_i|Z_i) = E(\beta_0 + \beta_1 CadreStatus_i + U_i|Z_i)$$

$$= \beta_0 + \beta_1 E(CadreStatus_i|Z_i) + E(U_i|Z_i)$$

$$= \beta_0 + \beta_1 \widehat{CadreStatus}_i$$

As such, our causal parameter of interest β_1 can be obtained from 2SLS. The example shown above greatly simplifies realistic models used in practice. However, the method of 2SLS is valid without loss of generality to more complicated models.

To interpret and analyze the accuracy of both the OLS and 2SLS results, we make an adjustment to its standard error using White's heteroscedasticity robust standard error. Such an adjustment will ensure the significance levels of the results are consistent when the variability of the variable cadre status is uneven across its range of values (i.e. heteroscedasticity) (White, 1980).

For an a posteriori test of IV validity, we test that the instrumental variables used are both relevant and exogenous. For relevance, we use the Wald test statistic to test that at least one variable in the set of IVs is relevant in explaining the change in the cadre status of an individual (Wald, 1943). Next, we use the Sargan-Hansen over-identifying restrictions test to offer a partial test of exogeneity. I.e., we test if the instrumental variables do not correlate with any variable other than cadre status. Note that this is a partial test of exogeneity, as there is an underlying assumption that there are enough instruments to allow us to identify the causal equation (Sargan, 1958; Hansen, 1982).

3.2 Data and Research Design:

The empirical data used here is from the “Three-City Survey” conducted in 1999 in the cities of Wuhan, Shanghai, and Xi’an. A total of 4,444 individuals were surveyed on questions regarding their social, economical, and political status (Xie and Pan, 2010). Out of the sample surveyed, of particular interest to us is a subset of 2,779 individuals that do not have missing values in our variables of interest.

For our regression models, we follow the convention prevalent in the literature. A log-linear model is used to study the percentage changes in household income from changes in the explanatory variables. Consistent with Xie et al. (2009), we introduce control variables that account for exogenous variations in each household’s city of residence, gender, education, and experience. Education is defined to be the highest education level attended (from 0 being no education to 5 being college graduate) and experience is the number of years of work experience (Xie and Pan, 2010).

For the response variable, the survey defines it as monthly household income in 1998 (Xie and Pan, 2010). Our causal explanatory variable of interest is the individual’s current cadre status. For the 2SLS regression, the individual’s first cadre status and the cadre status of the individual’s offspring is used as instrumental variables. The individual’s offspring cadre status is a binary variable with 1 indicating a cadre and 2 indicating otherwise, while the individual’s cadre status is ranked from 1 to 5 according to Table 1.

Table 1: Individual Cadre Status

Value	*Label*
1	Below-section Chief (below ke ji)
2	Section Chief (ke ji)
3	Department Head (chu ji)
4	Head of a Bureau or Higher (ju ji or above)
5	Not a Cadre
-99 (M)	Blank when response expected
-98 (M)	Don't know
-87 (M)	Not applicable: never worked

Note: Reprinted from *Study of Family Life in Urban China, 1999: Codebook* (p. 60) by Y. Xie and Z. Pan, 2010, Inter-university Consortium for Political and Social Research. Copyright 2010 by the Inter-university Consortium for Political and Social Research

An a priori argument on the validity of the two instrumental variables will be made here. A statistical test on their validity will be made a posteriori in Section 3.4. For relevance, it is arguable that the individual's first cadre status and their offspring's cadre status will be correlated with the individual's current cadre status. Such connections can stem from intergenerational political capital accumulation. For exogeneity, the individual's first cadre status nor their offspring's cadre status should have a material correlation with what the household earns at present.[2] As such, it is heuristically arguable that the two instrumental variables chosen are valid.

3.3 Results

Both an OLS and 2SLS regression are conducted and are of the functional form:

$$\begin{aligned} E(HouseholdIncome_i | CadreStatus_i) \\ &= \beta_0 + \beta_1 City_i + \beta_2 Gender_i + \beta_3 Education_i + \beta_4 Experience_i \\ &+ \beta_5 CadreStatus_i \end{aligned}$$

We note that for the 2SLS regression, we condition of the instrumental variables instead of the individual's cadre status. The R code for the two regressions is available under A.1. The heteroscedastic robust results for the OLS regression is displayed below. Note that this is a purely statistical interpretation.

[2] The offspring's financial status is surveyed separately. Thus, household income should exclude any contributions from adult offspring.

Table 2: OLS Log Earnings Regression

	Dependent variable:
	Log Monthly Household Income
Intercept	7.103***
	(0.089)
City	−0.253***
	(0.016)
Gender	0.065**
	(0.025)
Education	0.158***
	(0.009)
Experience	0.008***
	(0.001)
Cadre Status	−0.047***
	(0.009)
Observations	2,779
R^2	0.188
Adjusted R^2	0.187
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 2 confirms the statistical correlations observed in Xie et al. (2009) and Xie and Jin (2015). In particular, every rank increase in the individual's cadre status is correlated with, on average, a 4.7% increase in the household's income, holding all else constant.[3][4] This statistical correlation is statistically significant at all conventional significance levels, and is reasonably within the vicinity found by Xie et al. (2009) of 18.5%.[5]

We now move on to a causal analysis. The 2SLS regression model results is given in Table 3. We observe that for every rank increase in the individual's cadre status, it causes, on average, a 7.2% increase in the household's income, holding all else constant. Note the language used above; we mention that it causes, not that it is correlated with. This is due to the 2SLS regression model accounting for potential problems that may invalidate the causal interpretation of OLS.

[3] Note here that an increase in an individual's cadre rank decreases the $CadreStatus_i$ variable.
[4] This stems from the interpretation of log-linear regressions (Stock and Watson, 2015, pp. 315–324).
[5] Xie et al. (2009) uses a binary variable to characterize cadre status, whereas we use a 5-point scale for the variable. Therefore, it is reasonable for the result in Xie et al. (2009) to be approximately 5 times as large.

Table 3: 2SLS Log Earnings Regression

	Dependent variable:
	Log Monthly Household Income
Intercept	7.234***
	(0.121)
City	−0.256***
	(0.016)
Gender	0.071***
	(0.025)
Education	0.152***
	(0.010)
Experience	0.007***
	(0.001)
Cadre Status	−0.072***
	(0.017)
Observations	2,779
R^2	0.186
Adjusted R^2	0.185
Note:	*p<0.1; **p<0.05; ***p<0.01

From Table 3, after accounting for heteroscedasticity, the cadre status variable is statistically significant at all conventional significance levels. The magnitude is also comparatively larger than the statistical OLS findings, suggesting that the causal impact of the individual's cadre status on household income is more consequential than what previous statistical methods find in the literature. Such findings confirm our hypothesis; the statistical correlation found is indeed causal. It is not the byproduct of other underlying data generating processes that are concealed due to misspecifications arising from problems such as omitted variable bias and/or simultaneous causality.

3.4 Validity Check

3.4.1 Relevance

To check a posteriori that the instrumental variables are relevant, we use the Wald test statistic to test that at least one variable in the set of IVs is relevant in explaining the change in the cadre status of an individual (Wald, 1943). The null hypothesis is that both IVs are irrelevant, while the alternate hypothesis is at least one IV is relevant. The Wald test statistic follows a Chi-squared distribution with the degrees of freedom equal to the number of restrictions tested (two in this instance). The regression and Wald test results are displayed in Tables 4 and 5.

Table 4: First Stage 2SLS Relevance Regression

	Dependent variable:
	Cadre Status
Intercept	1.646***
	(0.173)
City	−0.077***
	(0.024)
Education	−0.100***
	(0.014)
Gender	0.220***
	(0.039)
Experience	−0.015***
	(0.002)
Offspring Cadre Status	0.110**
	(0.050)
First Cadre Status	0.646***
	(0.017)
Observations	2,779
R^2	0.439
Adjusted R^2	0.438
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5: Wald Test Results

Statistic	Degrees of Freedom	Chisq	Pr(>Chisq)
Value	2	753.986	0

From the Wald test in Table 5, it can be seen that the null hypothesis is convincingly rejected. Therefore, this indicates that the instrumentals are relevant in picking up changes in the endogenous variable $CadreStatus_i$.

3.4.2 Exogeneity

We now use the Sargan-Hansen over-identifying restrictions test to offer a partial test of exogeneity. A rejection of the null hypothesis indicates that the instrumental variables are endogenous (i.e., the IVs are correlated with variables other than $CadreStatus_i$), while a non-rejection is partial evidence supporting the exogeneity of the instrumental variables (Sargan, 1958;

Hansen, 1982). The degrees of freedoms here is the number of over-identified parameters (one in this case).[6] Table 6 displays the test results.

Table 6: Sargan-Hansen Over-Identifying Restrictions Test Results

Statistic	Degrees of Freedom	Chisq	Pr(>Chisq)
Value	1	3.5344	0.0601

From Table 6, we see that the test statistic is not rejected at conventional significance levels. Thus, we do not reject the null hypothesis and interpret it as evidence supporting exogeneity. However, the reader should caution that this is only a partial test of exogeneity, and that the test assumes that there are enough instruments to allow us to identify the causal equation.

From Sections 3.4.1 and 3.4.2, we can conclude that the instrumental variables used are both relevant and exogenous. As such, our choice of the instrumental variables is valid, and is appropriate to use in the first stage of the 2SLS regression model.

4. Limitations

Despite the numerous insights the 2SLS model brings to causality analysis, limitations of the model are inevitable. Flowing naturally from the previous section, the over-identifying restrictions test results in Table 6 are only marginally insignificant. Therefore, there is some possibility that a type 2 error exists. I.e., we are erroneously not rejecting the null hypothesis when it is false, and the instrumental variables are indeed correlated with other variables. As such, the reader should turn to theory to accept or reject the instrumental variables as exogenous or not.

[6] For a full explanation of the over-identifying restrictions test, please refer to Sargan (1958) and Hansen (1982).

Another limitation concerns the sample size of the "Three-City Survey" and the unavailability of any broader survey that instrumental variables can be readily found. The usable sample size is limited to 2,779 individuals. As such, while the 2SLS model is unbiased asymptotically, it is not guaranteed to unbiasedly approximate the true population causal parameter in finite samples (Stock and Watson, 2015, pp. 513–515). Additionally, as the name suggests, the "Three-City Survey" is only limited to three cities – Wuhan, Shanghai, and Xi'an, in which all can be considered to be politically-established (Xie and Pan, 2010). The geographic limitations of the sample may induce sample biases that cannot be resolved by the 2SLS model; namely, the external validity of our causal findings, and if the results found generalizes to the general population (Stock and Watson, 2015, pp. 378–385).

5. Conclusion

This paper introduces a novel addition to the literature regarding the causal relationship between household income and cadre status. Though a 2SLS framework, this paper overcomes many problems arising from standard OLS approaches regarding the causal interpretability of regression parameters. Using data from the "Three-City Survey", the authors find that the statistical correlation observed in the literature is indeed causal. In politically-established cities, an individual's cadre status causes positive rents to be incurred in household wealth accumulation. The magnitude of this causality relationship is also more pronounced than what conventional statistical methods observe; and is statistically significant at all conventional significance levels. Such causal information would be helpful to policy makers and academics in analyzing the impulse-response effects of policy changes and structural changes in society. It would also be valuable information to households on how to optimally allocate resources.

For further research, the extension of this causality analysis to China in aggregate would prove to be a valuable contribution to the literature. This may involve finding proxies for political capital in large scale national surveys such as the China Family Panel Studies (CPFS) and using a causality model such as the 2SLS model or the fixed-effects panel regression model to analyze the causality impacts of an individual's cadre status.[7] Similarly, adjustments to resolve the under-sampling biases for the ultra-rich may also be a constructive contribution to the literature. The suggestions above and finding methods to overcome the limitations in Section 4 would prove invaluable to the current literature; offering new insights into the causality relationship between an individual's cadre status and their household's wealth accumulation in contemporary China.

[7] An analysis on the merits of fixed-effects panel regression models can be found in Chapter 10 of Stock and Watson (2015).

References

Fu, B. (2021). *From Reds to Riches: A Literature Review on the Beneficiaries of Deng's Social-Economic Policies Post-1992*. Unpublished manuscript, the University of Melbourne.

Hansen, L. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, *50*(4), 1029-1054. https://doi.org/10.2307/1912775

Nee, V., & Opper, S. (2010). Political Capital in a Market Economy. *Social Forces*, *88*(5), 2105-2132. https://doi.org/10.1353/sof.2010.0039

Sargan, J. (1958). The Estimation of Economic Relationships using Instrumental Variables. *Econometrica*, *26*(3), 393-415. https://doi.org/10.2307/1907619

Stock, J., & Watson, M. (2015). *Introduction to Econometrics* (3rd ed.). Pearson.

Wald, A. (1943). Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large. *Transactions of the American Mathematical Society*, *54*(3), 426-482. https://doi.org/10.1090/s0002-9947-1943-0012401-3

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, *48*(4), 817-838. https://doi.org/10.2307/1912934

Xie, Y., & Jin, Y. (2015). Household Wealth in China. *Chinese Sociological Review*, *47*(3), 203-229. https://doi.org/10.1080/21620555.2015.1032158

Xie, Y., & Pan, Z. (2010). Study of Family Life in Urban China, 1999. *ICPSR Data Holdings*. https://doi.org/10.3886/icpsr28143.v1

Xie, Y., Lai, Q., & Wu, X. (2009). DANWEI AND SOCIAL INEQUALITY IN CONTEMPORARY URBAN CHINA. *Research in the Sociology of Work, 19*, 283-306. https://doi.org/10.1108/s0277-2833(2009)0000019013

Appendices:

A.1. R Code for Empirical Results

```r
library(haven)
library(stargazer)
library(AER)

# Data read
import_data = read_dta("28143-0001-Data.dta")

# Importing parameters of interest
CITY = import_data["CITY"]
SEX = import_data["SEX"]

EDU_LEVEL = import_data["HED_C7"]
EXP = import_data["EXP"]
HHINC = import_data["HHINC"]

CADRE1 = import_data["CADRE1"]
CADRE2 = import_data["CADRE2"]
CADRE_OFFSPRING = import_data["RCADREB"]

# Dataset aggregation
interest_ds = data.frame(CITY, SEX, EDU_LEVEL, EXP, HHINC, CADRE1, CADRE2, CADRE_OFFSPRING)

# Removes entries where observations are missing
remove_index = c()

for (index_entry in 1:nrow(interest_ds)) {
  entry = interest_ds[index_entry, ]
  entry_remove = FALSE

  # Logical tests to find missing values
  if (entry["EXP"] < 0) {
    entry_remove = TRUE
  }

  # The leq sign is to remove any values where the person was unemployed as it would
  # cause the log-linear regression to fail
  if (entry["HHINC"] <= 0) {
    entry_remove = TRUE
  }

  if (entry["CADRE1"] < 0) {
    entry_remove = TRUE
  }

  if (entry["CADRE2"] < 0) {
    entry_remove = TRUE
  }

  if (entry["RCADREB"] < 0) {
    entry_remove = TRUE
  }

  if (entry_remove == TRUE) {
    remove_index = c(remove_index, index_entry)
  }
}

# Adjusted aggregated dataset with no missing values
adj_interest_ds = interest_ds[-remove_index, ]

# OLS regression
ols_reg = lm(log(HHINC)~CITY+SEX+HED_C7+EXP+CADRE2, data=adj_interest_ds)
ols_reg_hc = coeftest(ols_reg, vcov=vcovHC)

# Table output
```

```
control_var_names = c("Intercept", "City", "Gender", "Education", "Experience",
                      "Cadre Status")

stargazer(ols_reg, title = "OLS Log Earnings (non-HC)",
          covariate.labels = control_var_names,
          dep.var.labels = "Log Monthly Household Income",
          intercept.bottom = FALSE, intercept.top = TRUE,
          omit.stat=c("f", "ser"))

stargazer(ols_reg_hc, title = "OLS Log Earnings (HC)",
          covariate.labels = control_var_names,
          dep.var.labels = "Log Monthly Household Income",
          intercept.bottom = FALSE, intercept.top = TRUE,
          omit.stat=c("f", "ser"))

# 2SLS regression
iv_reg = ivreg(log(HHINC)~CITY+SEX+HED_C7+EXP+CADRE2 |
                 CITY+SEX+HED_C7+EXP+RCADREB+CADRE1, data=adj_interest_ds)
iv_reg_hc = coeftest(iv_reg, vcov=vcovHC)

# Table output
control_var_names = c("Intercept", "City", "Gender", "Education", "Experience",
                      "Cadre Status")

stargazer(iv_reg, title = "2SLS Log Earnings (non-HC)",
          covariate.labels = control_var_names,
          dep.var.labels = "Log Monthly Household Income",
          intercept.bottom = FALSE, intercept.top = TRUE,
          omit.stat=c("f", "ser"))

stargazer(iv_reg_hc, title = "2SLS Log Earnings (HC)",
          covariate.labels = control_var_names,
          dep.var.labels = "Log Monthly Household Income",
          intercept.bottom = FALSE, intercept.top = TRUE,
          omit.stat=c("f", "ser"))

# Check for IV relevance
rel_reg = lm(CADRE2~CITY+HED_C7+SEX+EXP+RCADREB+CADRE1, data=adj_interest_ds)
rel_test = linearHypothesis(rel_reg, c("RCADREB = 0", "CADRE1 = 0"), test="Chisq",
                            vcov=vcovHC(rel_reg))

# Table output
control_var_names = c("Intercept", "City", "Education", "Gender", "Experience",
                      "Offspring Cadre Status", "First Cadre Status")

stargazer(rel_reg, title = "First Stage 2SLS Relevance Regression",
          covariate.labels = control_var_names,
          dep.var.labels = "Cadre Status",
          intercept.bottom = FALSE, intercept.top = TRUE,
          omit.stat=c("f", "ser"))

W = rel_test[[3]][2]
ProbW = pchisq(W, df=2, lower.tail=FALSE)
round(cbind(W, ProbW), 4)

# Check for IV exogeneity
residuals_iv_reg = iv_reg$residuals
overid_reg = lm(residuals_iv_reg~CITY+SEX+HED_C7+EXP+RCADREB+CADRE1,
                data=adj_interest_ds)

overid_test = linearHypothesis(overid_reg, c("RCADREB = 0", "CADRE1 = 0"), test="Chisq",
                               vcov=vcovHC(overid_reg))

# Degrees of freedom adjustment
J = overid_test[[3]][2]
ProbJ = pchisq(J, df=1, lower.tail=FALSE)

# Table output
round(cbind(J, ProbJ), 4)
```