

Name of the Participating Students: Wanqing Emily Su

High School: Tsinghua International School

Province: Beijing

Country/Region: China/North Division

Name of Advisor(s): Xia Wang

Affiliation of Advisor(s): Renmin University of China

Title of Research Report: Pension, Labor Supply and Moral

Hazard: Evidence from China's New Rural Pension Scheme

Pension, Labor Supply and Moral Hazard: Evidence from China's New Rural Pension Scheme*

Wanqing Emily Su^a

^aTsinghua International School, Beijing, China

September 15, 2025

Abstract

This study examines the impact of China's New Rural Pension Scheme (NRPS) on the labor supply and health-related risk behaviors of the rural elderly. Using panel data from the China Health and Retirement Longitudinal Survey (CHARLS, 2011–2018) and robust econometric methods, including two-way fixed effects, staggered difference-in-differences, and double machine learning, we find that NRPS receipt significantly reduces agricultural labor participation by approximately 10 percentage points, while its effects on non-agricultural work are smaller and less robust. This disparity is likely attributable to intra-household labor division patterns. We also document evidence of moral hazard: men with agricultural hukou increase their drinking frequency in anticipation of benefit eligibility. While the NRPS provides crucial financial security, our results suggest its effectiveness could be enhanced with complementary targeted public health interventions.

Key words: Labor Supply, Moral Hazard, New Rural Pension Scheme, Staggered DID, Double Machine Learning

*Address correspondence to: Wanqing Emily Su, Tsinghua International School, Zhongguancun North St. Haidian District, Beijing, PRC 100084, China; E-mail: EmilySuWanqing@gmail.com.

Table of Contents

1 Introduction 4

2 Background on New Rural Pension Scheme 8

3 Data 10

3.1 Data Description 10

3.2 Summary Statistics 13

4 Empirical Results 14

4.1 Baseline DID Estimation 15

4.2 Callaway and Sant’Anna’s Methods 17

5 Robustness Analyses and Tests 19

5.1 ATT by Groups and Calendar Time 20

5.2 Alternative Treatment Variables 22

5.3 Alternative Control Group 23

5.4 Test for the Overlap Condition 24

5.5 The Results for a Shorter Panel 25

5.6 Robustness Analysis Using Panels with Restricted Age Ranges 26

5.7 Double Machine Learning with Fixed Effects 27

5.8 Test of Moral Hazard 30

6 Heterogeneity Analysis 31

7 Conclusion 34

References 38

1 Introduction

Over the past four decades, China has achieved the world's largest scale of poverty alleviation and most rapid economic growth, a success largely attributable to its transformative economic reforms and opening-up policies (Qian, 2017). Yet, concurrently, the country has undergone one of the most profound demographic transitions globally, shifting from one of the fastest-growing populations to a nation with one of the lowest total fertility rates among developing countries. This demographic shift has precipitated rapid population aging, posing the central challenge of “*growing old before getting rich*.” As noted by Meng and Tuo (2022), this phenomenon exacerbates the national dependency burden and manifests in significant disparities across regions, industries, and between urban and rural areas.

A further consequence of China's profound economic transformation is its pronounced urban-rural disparity. The income gap between urban and rural areas has continued to widen in the twenty-first century, driven by differentials in human capital investment, access to economic opportunities, and the provision of essential services ranging from finance to social security. A critical aspect of this inequality is the unequal access to health insurance and social protection programs, which exposes rural residents to markedly different, and often more binding, economic constraints, thereby resulting in substantial welfare disparities.

To alleviate rural poverty and safeguard the basic livelihood of the elderly, the Chinese government launched a pilot of the New Rural Pension Scheme (NRPS) in 2009. According to the State Council (2009), this system integrates “individual contributions, collective assistance, and government subsidies, while combining social pooling with individual accounts.” By 2013, the scheme had achieved full coverage across all county-level administrative units. Under the NRPS, insured individuals become eligible to receive a fixed monthly pension upon reaching

age 60. This feature makes pension receipt a useful natural experiment for examining its impact on the labor supply (Ning et al., 2016), occupational choices, and health-related risk behaviors (Cheng et al., 2018) of the rural elderly.

Theoretically, pension income affects labor supply through two channels: an *income effect* and a *substitution effect*. The former allows older adults to work less by providing non-work income, while the latter makes leisure more attractive by effectively lowering the wage for continued work. Empirical literature has documented that pensions generally reduce labor supply in both developed and developing countries (de Carvalho Filho, 2008), suggesting that the income effect typically dominates. For non-means-tested programs like the NRPS, where benefits are not contingent on retirement, the substitution effect should theoretically be negligible, as the scheme does not alter the opportunity cost of leisure.¹ However, existing studies on the NRPS have produced heterogeneous estimates of its impact on labor supply; see, e.g., Ning et al. (2016) and Huang and Zhang (2021). To address this inconsistency, this paper employs a longer panel dataset to provide a more robust assessment of how these cash transfers shape the work decisions of older adults.

Generally speaking, social pension programs exert comprehensive influences on individual and household behavior. Regarding human capital investment, Tang et al. (2021) utilized data from the China Family Panel Studies (CFPS) and a propensity score matching difference-in-differences (PSM-DID) approach to demonstrate that the NRPS significantly increased educational expenditures for children aged 0–16 by 33.44 percent, with an even larger effect (45.29 percent) among school-aged samples. This underscores the significant spillover effects of non-educational social security policies on human capital accumulation.

The evidence on elderly labor supply, however, is more mixed. Several studies documented

¹A non-means-tested pension program is a retirement benefit paid to individuals based solely on their age and/or prior contributions, without any consideration of their current income or wealth.

a reduction in labor participation. [Huang and Zhang \(2021\)](#), using CHARLS and CFPS data within a DID framework, found that the NRPS reduced labor supply among rural residents aged 60 and above by 3 percentage points, with agricultural labor participation falling by 3.6 percentage points, while also improving health outcomes. Similarly, [Huang et al. \(2014\)](#) reported that each additional yuan of NRPS pension income reduced annual labor time by approximately 0.01 days, with a stronger suppressing effect on agricultural work. Their study also identified heterogeneous effects by gender (stronger for men), age (more pronounced near age 60), and region (larger in the West than the East/Central regions). [Zhang et al. \(2015\)](#) further confirmed that the NRPS increased the personal income of rural elderly individuals, particularly non-labor income, reduced household poverty incidence by up to 11 percentage points under the old poverty line, and modestly reduced the supply of labor by increasing the propensity to retire. Furthermore, the pension scheme modestly stimulated household consumption, enhanced subjective well-being, as evidenced by reduced depressive symptoms and greater life satisfaction. In contrast, [Ning et al. \(2016\)](#), employing a regression discontinuity DID (RD-DID) design, found no significant negative effect of NRPS on overall labor force participation among the rural elderly, except for a slight reduction among those in poor health. The authors attribute this null result to the program's modest benefit levels and potential crowding-out of inter-generational transfers.

Finally, while not focused on pensions, the work of [Fu et al. \(2017\)](#) on the New Rural Cooperative Medical Scheme (NCMS) provides a relevant parallel. They documented ex-ante moral hazard, wherein healthier individuals reduced preventive health behaviors, highlighting how behavioral responses to social security programs can vary significantly across subgroups. This type of heterogeneity could also help explain the mixed findings on the NRPS.

This paper investigates the impact of pension receipt on labor supply and health-related

risk behaviors among rural residents in China. We utilize four waves of data from the China Health and Retirement Longitudinal Survey (CHARLS), which provides a larger sample size and a longer panel period than those used in prior studies, such as [Ning et al. \(2016\)](#). This extended temporal dimension not only offers greater statistical power to identify the long-term effects of the pension program on various outcomes but also facilitates a more comprehensive set of robustness checks. Methodologically, this study advances upon the existing literature by employing a suite of robust econometric techniques. Our estimation strategy includes two-way fixed effects (TWFE), aggregated and separate estimators using the recent causal inference methods for staggered adoption (CSDID), comprising event study, group study, and calendar time study approaches, as well as double machine learning methods. This multi-faceted approach allows for a more credible and nuanced comparison of estimates.

Our analysis reveals a consistently significant negative effect of receiving NRPS benefits on agricultural labor supply. In contrast, the effect on non-agricultural labor supply is smaller and less robust, showing variation across model specifications. We hypothesize that this divergence is driven by intra-household division of labor patterns in rural areas, a mechanism supported by our heterogeneous effect analyses.

Furthermore, this study examines potential behavioral changes resulting from increased leisure time, an aspect often overlooked in the literature. We identify an intriguing anticipation effect: males with agricultural *hukou* are more likely to increase their drinking frequency in the period immediately preceding pension eligibility. This finding suggests that although pension programs effectively provide financial security and leisure, they may also introduce moral hazard, potentially encouraging risky behaviors. Consequently, our results indicate that policymakers should consider complementing such programs with targeted health interventions to mitigate these unintended consequences.

The remainder of this paper is organized as follows. Section 2 gives a short introduction on the NRPS in China. Section 3 describes the data. The empirical strategy and results are provided in section 4. Section 5 conducts some robustness checks and tests about moral hazard. Section 6 presents a heterogeneous analysis. Section 7 concludes.

2 Background on New Rural Pension Scheme

The New Rural Pension Scheme (NRPS) is a pension program designed specifically for rural *hukou* holders. It operates as a pay-as-you-go system, financed primarily by the central government and supplemented by local government subsidies. Rural residents can enroll voluntarily, beginning contributions at age 45 and starting to receive pension benefits at age 60. If an individual's contribution period is less than 15 years upon reaching age 60, a lump-sum payment is required to cover the deficit. In the initial 2009 policy, individuals aged 60 or above in 2009 became eligible for benefits without making contributions, provided their adult children were enrolled in the scheme. This family binding requirement was later abolished.

Participants can choose their annual contribution tier, typically ranging from 100 to 500 yuan, though some affluent coastal regions permit higher tiers (e.g., 2,500 yuan per year). Local governments provide a minimum annual subsidy of 30 yuan per participant, irrespective of the chosen tier. The central government fully subsidizes the basic pension in central and western regions, while covering 50% of the cost in eastern regions.

Before 2014, the minimum monthly pension was 55 yuan. Although modest in absolute terms, this amounted to roughly one-quarter of the median monthly income of the rural elderly, estimated by Huang and Zhang (2021), and thus had considerable welfare effects. The basic pension was increased to 75 yuan per month nationally in 2014, with provisions for future adjustments.

The labor supply decisions of the elderly are shaped by the NRPS through the standard economic lenses of *income* and *substitution* effects. First, the pension generates an income effect: by providing non-labor income, it relaxes budget constraints and reduces the need to work to maintain consumption. Evidence from [Huang and Zhang \(2021\)](#) supports this, showing that the NRPS increased household income by 18% and raised food expenditures by 9.6%, while simultaneously decreasing overall labor force participation by 3.0 percentage points. Second, the scheme's specific rules can create a substitution effect by changing the opportunity cost of leisure. The design of contribution requirements and benefit calculations alters the relative price of labor versus leisure and thereby indirectly encourages or discourages work. For example, if pension rules were designed such that continued work beyond the eligibility age increased the amount of benefits or accrued interest, one could expect this to encourage participants to remain in the labor force. Conversely, if work income or pension contributions beyond the eligibility age were no longer eligible for tax or subsidy benefits, individuals might instead be incentivized to reduce labor supply.

A critical institutional feature of the NRPS is that it is not *means-tested*. Pension receipt is not conditional on retirement or meeting any other strict eligibility thresholds beyond age and contribution history. Thus, the program does not impose any additional restrictions on receiving pension payments once enrolled participants reach age 60. Consequently, this design implies that the estimated effects of the pension are likely driven primarily by a pure income effect, resulting from the direct financial transfer.

However, the NRPS may also distort labor supply decisions before payments begin. The program requires individuals to contribute for at least 15 years to qualify for a full pension at age 60, which often pressures those who enroll in their late 40s or 50s, creating liquidity constraints until payments commence at age 60 ([Ning et al., 2016](#)). In rural areas, where adult

children traditionally provide substantial support, these public pension transfers may crowd out private filial support. If this reduction is not offset by the modest pension, older adults may remain in the labour force to maintain living standards. [Ning et al. \(2016\)](#) found that NRPS receipt did not reduce, and may have slightly increased, elderly labor supply. They speculated that the initial family-binding rule (which required adult children to contribute for their parents to receive the basic pension) could reduce transfers, but their test for this crowd-out effect was not statistically significant. They attributed this partly to data limitations, as the first wave of CHARLS included many individuals already over 60 when the NRPS launched. In contrast, [Huang and Zhang \(2021\)](#) propose a complementary explanation centered on labor reallocation for working-age participants. They argue that the need to make annual contributions (100–500 yuan) incentivizes individuals under 60 to shift from agricultural work to better-paid local wage labor, which provides the immediate cash needed to cover costs. Consequently, the observed reduction in agricultural labor is replaced by non-agricultural work, leading to no significant net improvement in economic or health welfare for the working-age cohort.

For further details on the institutional background, see [Ministry of Finance and Ministry of Human Resources and Social Security \(2011\)](#), [Ma and Zhou \(2014\)](#), [Ning et al. \(2016\)](#), and [Huang and Zhang \(2021\)](#).

3 Data

This section describes the dataset and presents summary statistics.

3.1 Data Description

This study utilizes data from the China Health and Retirement Longitudinal Survey (CHARLS), a nationally representative longitudinal survey of Chinese residents aged 45 and older conducted

by Peking University every two or three years (Zhao et al., 2014).² The dataset provides comprehensive information on the elderly population. To analyze the effect of the NRPS on labor supply and behavioral outcomes, variables were selected from four core modules: demographic background, health status and functioning, health care and insurance, and work, retirement, and pension. The analysis uses all waves prior to the COVID-19 pandemic, namely, 2011, 2013, 2015, and 2018, to avoid potential confounding effects from pandemic-related lockdowns and behavioral changes.³

The primary independent variable is an interaction term, $NRPS60_{it}$, constructed as $NRPS_{it} \cdot \mathbf{1}\{Age_{it} \geq 60\}$, where $\mathbf{1}\{\cdot\}$ is the indicator function and $NRPS_{it}$ is a dummy variable that takes value 1 if individual i participates in the NRPS at year t . This captures the local average treatment effect of being both enrolled in the NRPS and over the eligibility age of 60, following the identification strategy of Chang (2013), Zhang et al. (2015), and Ning et al. (2016). For robustness, an alternative treatment variable, OAP_{it} , is used, which directly indicates whether the respondent i received old-age pension payments from the NRPS at year t . The correlation between $NRPS60$ and OAP is high (0.9), supporting the use of OAP as a valid alternative measure.

The dependent variables include two labor supply measures: *Agricultural* and *Non-agricultural*, both of which are dummy variables. The former indicates whether the individual engaged in agricultural work for more than 10 days, while the latter indicates whether they worked for at least one hour in the previous week (capturing wage-earning work). To test for moral hazard associated with receiving a pension, smoking status and monthly drinking frequency are also used as outcomes, yielding the variables *Smoke* and *Drink*, respectively.

²The CHARLS project begins sampling from age 40 to account for future attrition; respondents are then followed up biennially, though some may drop out in subsequent waves.

³The 2020 wave was excluded as it did not collect pension and social security data, and the pandemic and the lock-down policy may have drastically altered economic behavior and labor market conditions.

Following [Fu et al. \(2017\)](#) and [Ning et al. \(2016\)](#), we select control variables based on their exogeneity and low missing rates. These include hukou status (*Hukou*), education level (*Edu*), marriage status (*Marriage*), pain level (*Pain*), availability of a helper (*Helper*), income (*Income*), self-rated health condition (*Health*), and enrollment in other resident pension programs (*Resident_pension*). Variable definitions are provided in Table 1.

Table 1: Variable Definitions

Variables	Definition
Agricultural	Dummy: 1 if engaged in agricultural work for more than 10 days; 0 otherwise.
Non-Agricultural	Dummy: 1 if worked (earn a wage, run your own business and unpaid family business work) for at least one hour in the previous week; 0 otherwise.
Smoke	Dummy: 1 if currently smokes; 0 otherwise.
Drink	Monthly drinking frequency: none (0), drink but less than once a month (1), drink more than once a month (2).
NRPS60	Interaction: 1 if participates in the NRPS and aged 60 years or older; 0 otherwise.
NRPS	Dummy: 1 if participates in the NRPS; 0 otherwise.
OAP	Dummy: 1 if receives NRPS old-age pension; 0 otherwise.
Age	Age in years (current year minus birth year).
Hukou	Hukou type: agricultural hukou (1), non-agricultural hukou (2), unified Residence hukou (3), do not have hukou (4).
Edu	Education level: from no formal education illiterate (1) to post-graduate, master's degree (10)
Marriage	Marital status: married with spouse present (1), married but not living with spouse (2), separated (3), divorced (4), widowed (5), never married (6), cohabitating (7). For Regression: Dummy variable: 1 if married with spouse present; 0 otherwise.
Gender	Dummy: 1 (male); 0 (female).
Pain	Dummy: 1 if currently has body pain; 0 otherwise.
Helper	Dummy: 1 if expects future help; 0 otherwise.
Income	Dummy: 1 if received wage/bonus last year; 0 otherwise.
Health	Dummy: 1 if self-rated health is "good" or better; 0 otherwise.
Resident_pension	Dummy: 1 if enrolled in other pension programs; 0 otherwise.

3.2 Summary Statistics

The data from the four waves were merged using individual IDs to construct a balanced panel. To ensure identification for causal analysis, individuals who were always treated (i.e., those received NRPS pension income for all four periods, thus exhibiting no variation in treatment status) were excluded from the sample. The final balanced panel consists of 3,396 individuals observed across all four waves, with ages ranging from 22 to 95.

Table 2 presents the summary statistics. As shown, females constitute 86% of the sample. This gender imbalance likely stems from two factors: a higher likelihood of male absence from the household due to migrant labor during survey periods and a potentially higher survey response rate among females. Since the analysis relies on a balanced panel, applying the sampling weights provided by CHARLS was deemed inappropriate. To address concerns regarding sample representativeness, robustness checks were conducted using an unbalanced three-wave panel and by restricting the age range to include more male respondents.

The summary statistics in Table 2 further reveal that approximately 60% of the sample engaged in agricultural work, while 13% participated in non-agricultural employment. This labor pattern is reflected in income sources, with only about 13% of respondents having received wages or a bonus in the previous year. Regarding pension coverage, nearly 80% of the elderly were enrolled in a resident pension program (including but not limited to the NRPS), while 20% specifically received the NRPS old-age pension. Finally, health-related risk behaviors were relatively uncommon in this predominantly female sample: the rate of smoking was close to zero, and the average monthly drinking frequency was not high.

This study introduces several refinements relative to the existing literature. First, health status is measured as the average of three self-reported health scores (on a 1–6 scale from very good to very poor), providing a more robust measure. Second, the variable *Resident_pension*

is included to control for the influence of other public pension schemes. Third, rather than controlling for regional fixed effects (e.g., east v.s. central China), we employ individual fixed effects in the panel data models to account for time-invariant individual heterogeneity, following the standard approach for fixed effects panel data analysis.

Table 2: Summary Statistics

Variables	N	Mean	SD	Min	Max
Agricultural	13,584	0.599	0.490	0	1
Non-Agricultural	13,584	0.129	0.335	0	1
Smoke	13,584	0.00272	0.0521	0	1
Drink	13,584	0.301	0.668	0	2
NRPS60	13,584	0.206	0.404	0	1
NRPS	13,584	0.460	0.498	0	1
OAP	13,584	0.188	0.391	0	1
Age	13,584	59.42	8.938	22	95
Hukou	13,584	1.193	0.418	1	4
Edu	13,584	3.108	1.926	1	10
Marriage	13,584	1.535	1.297	1	7
Gender	13,584	0.140	0.347	0	1
Pain	13,584	0.439	0.496	0	1
Helper	13,584	0.709	0.454	0	1
Income	13,584	0.134	0.340	0	1
Health	13,584	3.180	0.924	1	5
Resident_pension	13,584	0.782	0.413	0	1

4 Empirical Results

This section presents the empirical results, beginning with a baseline difference-in-differences (DID) estimation and then applying the method of [Callaway and Sant' Anna \(2021\)](#) to account for staggered treatment adoption.

4.1 Baseline DID Estimation

First, we consider the standard panel with two way fixed effects (TWFEs) to identify the causal effect:

$$Work_{it} = \beta \cdot NRPS60_{it} + \gamma' X_{it} + \lambda_i + \phi_t + \epsilon_{it} \quad (1)$$

where $Work_{it}$ denotes agricultural work or non-agricultural work, $NRPS60_{it}$ is the interaction term defined in the last section, X_{it} denotes a collection of the control variables, λ_i and ϕ_t are individual fixed effects and year fixed effects, respectively, and ϵ_{it} is the idiosyncratic error term. The core parameter of interest is β in Equation (1). We treat the model in (1) as a baseline model.

Table 3: Baseline DID Estimation

Variables	Type of Work			
	(1) Agricultural	(2) Agricultural	(3) Non-Agricultural	(4) Non-Agricultural
NRPS60	-0.024** (0.011)	-0.026** (0.011)	-0.003 (0.007)	-0.006 (0.007)
Pain		0.006 (0.008)		-0.018*** (0.006)
Helper		0.025*** (0.008)		-0.008 (0.006)
Income		-0.018 (0.012)		0.250*** (0.014)
Health		-0.005 (0.005)		-0.010*** (0.004)
Resident_pension		-0.019 (0.022)		0.035** (0.018)
Constant	0.604*** (0.002)	0.616*** (0.024)	0.130*** (0.001)	0.116*** (0.019)
Hukou, Edu, Marriage		✓		✓
TWFE	✓	✓	✓	✓
Observations	13,584	13,584	13,584	13,584
R-squared	0.669	0.671	0.570	0.603

Note: Robust standard errors are in parentheses, *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. Hukou, Edu and Marriage are dummy variables here.

Table 3 reports the estimation results for the baseline model, where columns (1) and (3) do not include any control variables, whereas columns (2) and (4) include the control variables described previously. In all specifications, both individual and time fixed effects are included. Since *Hukou*, *Edu* and *Marriage* are time-invariant variables that do not change over this short time span, their coefficients cannot be identified once the individual effect term λ_i is included and are therefore absorbed by it.

According to Table 3, the inclusion of control variables has only a slight effect on the magnitude of the coefficient for *NRPS60*. The results indicate that receiving the NRPS pension significantly reduces the probability of participating in agricultural work. In contrast, its effect on non-agricultural work is statistically insignificant. This disparity can potentially be explained by several factors. First, the majority of elderly individuals in the sample hold agricultural *hukou* and thus have limited engagement in wage-earning employment. Second, intra-household labor specialization in rural areas may play a role: women often specialize in agricultural activities, while men are more likely to seek off-farm wage work. This interpretation is consistent with the high proportion of female respondents in the dataset. The remainder of this section examines these conjectures through a heterogeneous effects analysis.

Furthermore, individuals with familial or social support (i.e., those with family members, relatives, or friends to care for them) exhibit a higher probability of participating in agricultural work. The magnitude of this effect is comparable to that of receiving pension payments. Turning to non-agricultural work, individuals reporting physical pain are significantly less likely to participate, a finding consistently supported by the significantly negative coefficient on *Health*, a composite measure of self-evaluated health status. Conversely, higher income levels and enrollment in other resident pension programs (a potential proxy for urban residence) are associated with a greater likelihood of engaging in non-agricultural work. This suggests that

urban residents and those living in county towns are primarily motivated by wage earnings to remain in the labor force. For these individuals, NRPS benefits likely constitute only a small portion of their total income and therefore have limited impact on their labor supply decisions, even if they are enrolled in the scheme.

4.2 Callaway and Sant’Anna’s Methods

Since this study features staggered treatment adoption, the traditional two-way fixed effects estimator (TWFE) is susceptible to bias from negative weights (De Chaisemartin and d’Haultfoeuille, 2020). This issue implies that the TWFE estimate is not a proper weighted average of treatment effects; it can overweight certain comparisons and, in extreme cases, even produce an estimate with the opposite sign of the true average treatment effect on the treated (ATT).

To address this issue, we employ the method proposed by Callaway and Sant’Anna (2021), which is specifically designed for DID estimation with staggered adoption (CSDID). This approach first identifies group-time average treatment effects $ATT(g, t)$ separately and then aggregates these parameters using appropriate weights. Here, each $ATT(g, t)$ is identified by comparing a given treatment group g to an appropriate control group consisting only of units that are not yet treated at time t . The overall effect is then aggregated using non-negative and interpretable weights. This approach ensures that the estimator is robust to treatment effect heterogeneity and avoids the negative weighting problem inherent in TWFE. The resulting estimator is particularly useful for examining effect heterogeneity, including dynamic responses based on the length of exposure to the treatment (event study), as well as differential effects across various treatment groups. It is worth noting that while other robust estimators have been proposed for staggered DID designs, e.g., Sun and Abraham (2021) and Borusyak et al. (2024), we adopt the CS estimator for its intuitive aggregation of group-time effects. Formally

comparing these methods remains a valuable avenue for future research.

The dynamic effects of receiving NRPS payments are presented in Figure 1 (agricultural work) and Figure 2 (non-agricultural work). These event-study plots estimate the ATT relative to the time of treatment adoption, using the never-treated group as the control and including the same set of control variables as in the baseline DID specification. The results indicate a significant reduction in labor supply, for both agricultural and non-agricultural work, following the receipt of NRPS payments. Furthermore, the pre-treatment ATT estimates are statistically insignificant, lending support to the parallel trends assumption underlying our research design. This also provides no evidence of anticipation effects prior to the actual receipt of the pension.

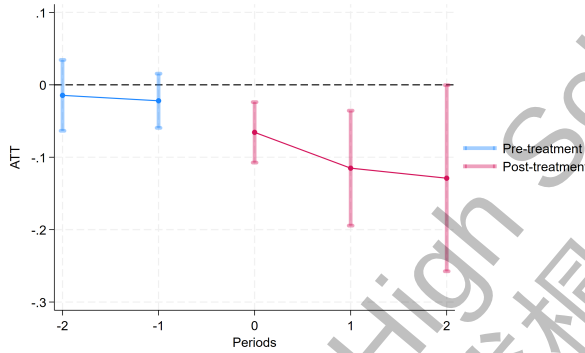


Figure 1: Event Study for Agricultural Work

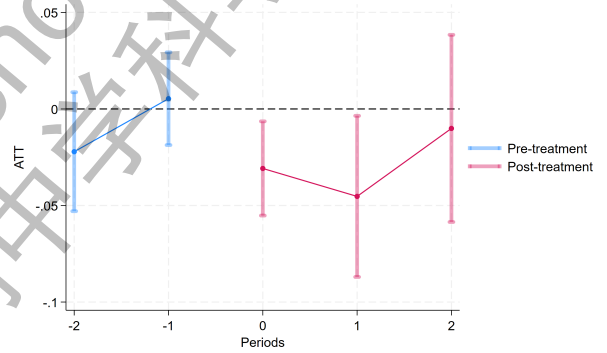


Figure 2: Event Study for Non-Agricultural Work

Table 4 presents the aggregated ATT across all treatment groups and time periods. These results are consistent with the event-study estimates, indicating that receiving NRPS reduces the probability of participating in agricultural work by approximately 10 percentage points and non-agricultural work by about 3 percentage points. The difference in effect magnitude between the two types of work may be attributed to several factors. First, it likely reflects the predominant *hukou* type among the elderly in the sample, which shapes their occupational access. Individuals with agricultural *hukou* typically have lower incomes. Consequently, the NRPS payment constitutes a larger share of their monthly income, leading to a stronger income

effect compared to those with non-agricultural *hukou* or those already engaged in wage-earning employment. Furthermore, agricultural work is generally more physically demanding than non-agricultural jobs. The pension income thus allows elderly individuals to preferentially reduce their engagement in the more strenuous agricultural activities.

Table 4: Aggregate ATT of CSDID Estimation

Variables	Type of Work			
	(1) Agricultural	(2) Agricultural	(3) Non-Agricultural	(4) Non-Agricultural
ATT	-0.050*** (0.013)	-0.097*** (0.031)	-0.019*** (0.008)	-0.031** (0.013)
Control Variables		✓		✓
Observations	13,584	13,584	13,584	13,584

Note: Robust standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Control variables are same as baseline DID, which include Pain, Helper, Income, Health, Resident, as well as Hukou, Education and Marriage dummy.

5 Robustness Analyses and Tests

This section presents a series of analyses that bolster our main findings, including an exploration of effect heterogeneity, robustness checks, tests of our identifying assumptions, results from alternative estimation strategies, and a test for moral hazard. First, we explore treatment effect heterogeneity by estimating the ATT for cohorts based on their initial treatment period or calendar time (Section 5.1). Next, we conduct a series of robustness checks on our aggregated result. In Section 5.2, we replace the key independent variable, *NRPS60*, with *OAP*. Section 5.3 considers alternative definitions of the control group. We then examine the overlap assumption required for the Callaway and Sant’ Anna (2021) estimator in Section 5.4. Subsequently, we assess robustness to sample composition using a shorter panel of only

three waves (Section 5.5) and two narrower age-restricted panels for individuals aged 50–70 and 55–65, respectively (Section 5.6). Finally, in Section 5.7, we employ an alternative estimation strategy, applying the double machine learning (DML) method to estimate the ATT, and we extend our analysis to perform a test for moral hazard in Section 5.8.

5.1 ATT by Groups and Calendar Time

The identification of distinct group-time average treatment effects enables the use of alternative weighting schemes to analyze effect heterogeneity across cohorts or over calendar time. Figures 3 and 4 show the ATT by group for agricultural work and non-agricultural work, respectively. In our study, individuals are classified into four treatment cohorts (groups) based on the calendar year in which they first became eligible for and began receiving the NRPS pension. In the figures, group G2 corresponds to the cohort first treated in the second wave (2011–2013), G3 to the cohort treated between 2013 and 2015, and G4 to the cohort treated between 2015 and 2018. When reporting the ATT by group, we track the outcome dynamics for each cohort relative to its own pre-treatment baseline period. The results, presented in Figures 3 and 4, show that the group-averaged ATT remains significant and negative for both agricultural work and non-agricultural work. However, when disaggregated by groups, the significant effects are more specific. For agricultural work, a significant negative effect is observed only in Group 2 (at the 5% level). For non-agricultural work, a significant effect is also found only in Group 2, though at the weaker 10% level. One potential explanation for this pattern is that the marginal treated units (i.e., the new additions) in the 2015 and 2018 cohorts are considerably smaller than the baseline group in 2013.⁴

In addition to analyzing effects by group, we also examine the evolution of the ATT over

⁴The number of treated units was 781 in 2013, compared to 166 new units in 2015 and 120 in 2018.

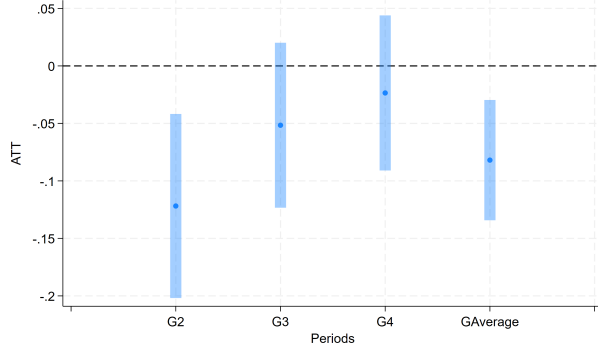


Figure 3: ATT by Group for Agricultural Work

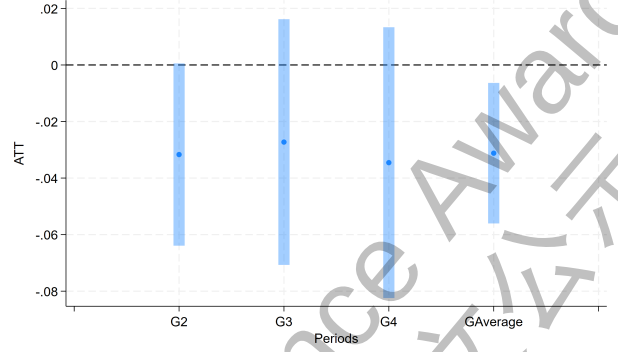


Figure 4: ATT by Group for Non-Agricultural Work

calendar time. For this analysis, we aggregate across groups to obtain the average treatment effect in a given survey wave, using individuals who have not yet been treated as controls.

Figures 5 and 6 present the ATT trends for engagement in agricultural and non-agricultural work, respectively. The calendar-time-averaged ATT is significantly negative for both sectors. However, while the ATT is negative and statistically significant in every time period for agricultural work, the effects for non-agricultural work are statistically indistinguishable from zero in all periods. This suggests that the income effect of the NRPS is more pronounced in agricultural work, both in terms of magnitude and statistical significance. This aligns with the earlier observation that, for individuals engaged in agriculture, the pension represents a more substantial portion of their overall wealth and income. Moreover, as shown in Figures 5 and 6, the treatment effect of pension receipt is persistent and shows no sign of decay beyond the initial treatment period: the average treatment effects on the treated (ATT) for T2, T3, and T4 (the second, third, and fourth periods after treatment) remain consistent in both size and statistical significance. This indicates that the impact of pension income is stable over time and represents a long-term effect, consistent with the policy's original intent. It is worth noting that such an analysis would not be feasible without a longer panel dataset, making this finding an important supplement to earlier studies such as [Ning et al. \(2016\)](#); [Huang et al. \(2014\)](#) and

Huang and Zhang (2021).

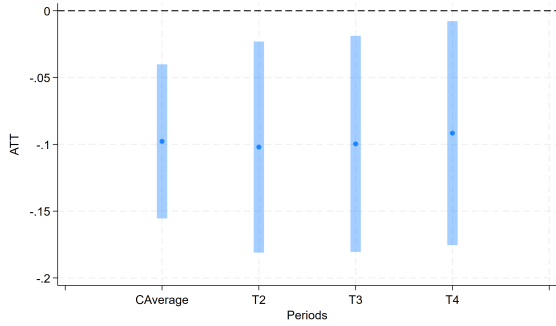


Figure 5: ATT by Calendar for Agricultural Work

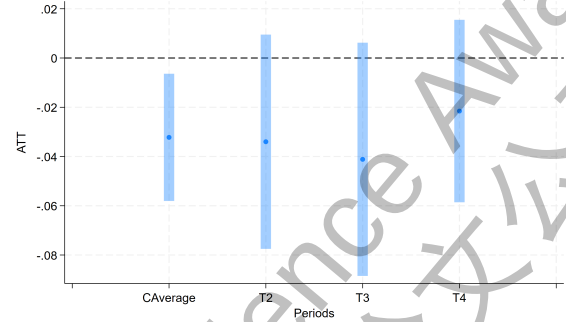


Figure 6: ATT by Calendar for Non-Agricultural Work

5.2 Alternative Treatment Variables

As mentioned in Section 3.1, *OAP* is an alternative, self-reported measure for NRPS participation, in contrast to the objective, data-derived *NRPS60* variable. To test the robustness of the main findings, we replicate the event study analysis using CSDID with the *OAP* variable. The results, presented in Figures 7 and 8, show that the effect on agricultural work participation remains consistent with the baseline analysis. For non-agricultural work, we find negative pre-treatment effects in the two periods preceding treatment, though the effects in all other periods remain unchanged. We caution against interpreting this as a violation of the parallel trends assumption, as the ATT is not significant in the first pre-treatment period. It is unclear why participants would reduce labor supply two periods before treatment due to anticipation but not in the period immediately prior. Moreover, these pre-treatment effects disappear when altering the measurement of the dependent variable, an change that does not affect the overall post-treatment results.

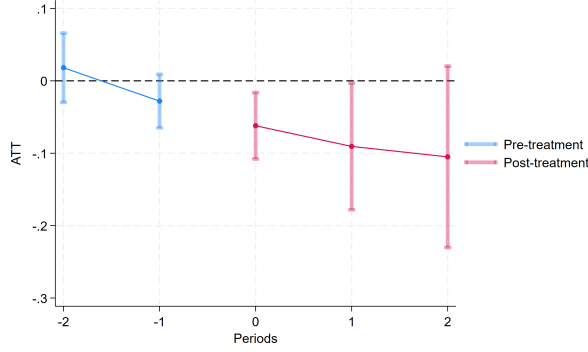


Figure 7: OAP and Agricultural Work

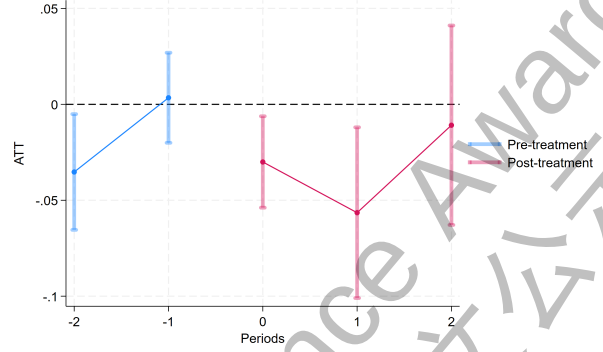


Figure 8: OAP and Non-Agricultural Work

5.3 Alternative Control Group

Since participation in the NRPS is voluntary, using “never-treated” units as the comparison group may be problematic due to potential selection bias. To address this concern and test the robustness of the results, we employ a control group comprising “not-yet-treated” units, as suggested by Callaway and Sant’Anna (2021). This approach relies on the assumption of conditional parallel trends between currently treated groups and those that have not yet been treated. Figures 9 and 10 present the dynamic ATT for agricultural and non-agricultural work, respectively. The results show that the estimated coefficients for the pre-treatment periods are statistically insignificant, supporting the validity of the parallel trends assumption, and that the post-treatment effect estimates are consistent with those derived using the “never-treated” control group.

This robustness check also suggests that the selected covariates are sufficient to control for observable differences between the treated and control groups. The consistency of results across alternative control group definitions indicates that, conditional on these covariates, the selection into NRPS participation is as good as random with respect to labor supply outcomes, thereby mitigating concerns about selection bias on unobservables.

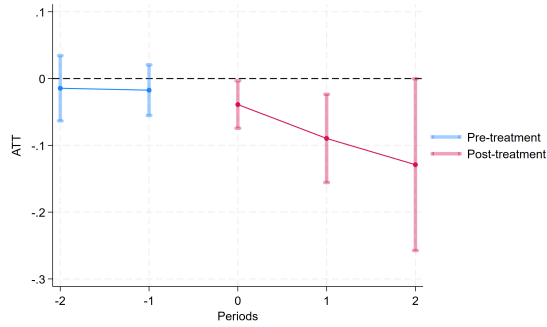


Figure 9: Not-yet-treated for Agricultural Work

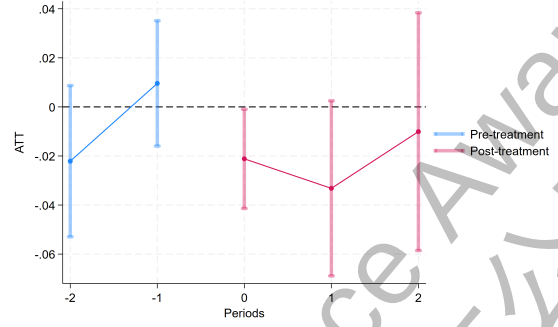


Figure 10: Not-yet-treated for Non-Agricultural Work

5.4 Test for the Overlap Condition

A key assumption of the [Callaway and Sant'Anna \(2021\)](#) estimator is the overlap condition, which requires that the estimated propensity score is bounded away from both zero and one. To assess this assumption, we plot the distribution of propensity scores for the treatment and control groups in [Figure 11](#). These scores were estimated using a logit model. The plot indicates that some units indeed have propensity scores very close to zero or one, which can be problematic as these observations may exert excessive influence and dominate the estimation results.

A standard approach to mitigate this issue is truncation, whereby units with propensity scores above or below specified thresholds are excluded from the analysis. To test the robustness of our findings, we truncate the sample by dropping observations with propensity scores above 0.9 or below 0.1. This process ensures common support and retains 2,487 units, representing approximately 73% of the original sample. The results obtained from this truncated sample, presented in [Figures 12 and 13](#), are largely consistent with the baseline estimates. The only notable difference is that the ATT for non-agricultural work in the first post-treatment period, which was previously significant only at the 10% level, is no longer significant in this robustness check.

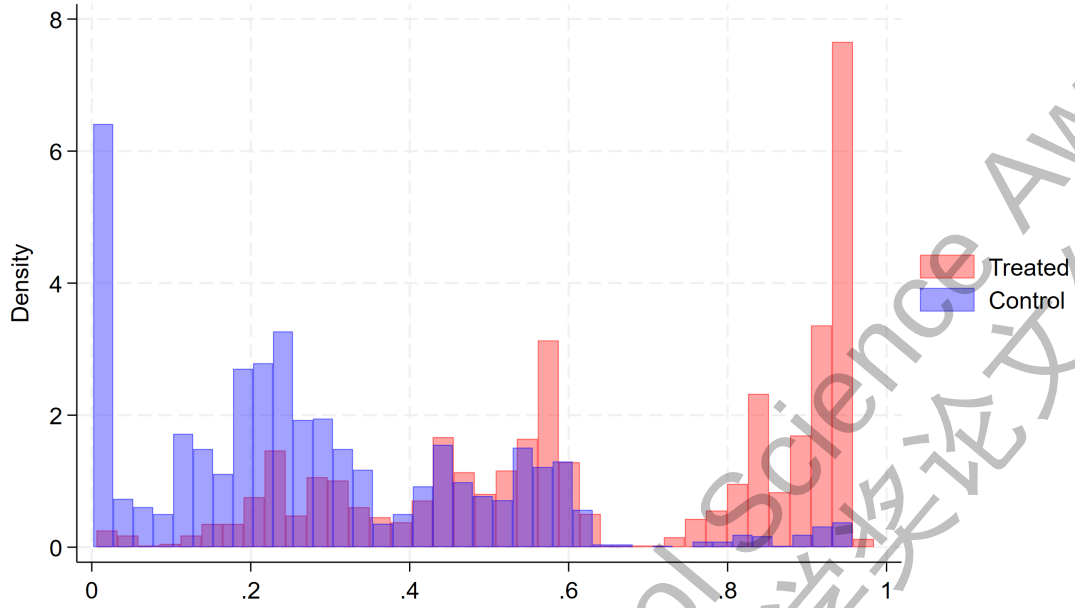


Figure 11: Propensity Score

5.5 The Results for a Shorter Panel

A further concern regarding the previous analyses stems from the high proportion of female respondents in the sample, which may affect the generalizability of the findings. To assess whether the results are sensitive to sample composition over time, we test their robustness by shortening the panel and removing the 2018 period. This choice is motivated by the finding in Section 5.1 that the ATTs are primarily driven by the 2013 treatment group. The results are presented in Table 5. Columns (1) and (2) use *NRPS60* as the treatment variable, while columns (3) and (4) use *OAP*. The dependent variables are agricultural work in columns (1) and (3) and non-agricultural work in columns (2) and (4). The results for non-agricultural work differ from the full panel; the coefficients for both *NRPS60* and *OAP* are no longer statistically significant. In contrast, the results for agricultural work remain significantly negative, confirming the robustness of the main findings.

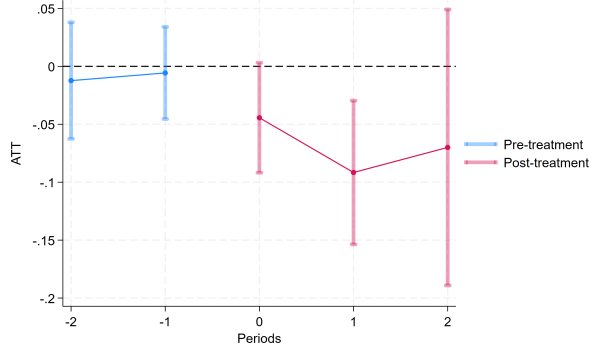


Figure 12: Truncated PS for Agricultural Work

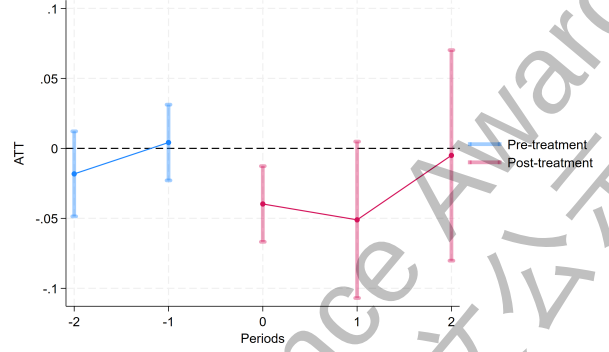


Figure 13: Truncated PS for Non-Agricultural Work

Table 5: The Results for a Shorter Panel

Variables	NRPS60		OAP	
	(1) Agricultural	(2) Non-Agricultural	(3) Agricultural	(4) Non-Agricultural
ATT	-0.055** (0.026)	-0.019 (0.014)	-0.052* (0.027)	-0.020 (0.013)
Control Variables	✓	✓	✓	✓
Observations	11,694	11,694	11,694	11,694

Note: Robust standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Control variables are same as before. All the results are estimated by CSDID.

5.6 Robustness Analysis Using Panels with Restricted Age Ranges

A potential concern is that the broad age range in the sample may bias the results. To mitigate this, we conduct a robustness check by restricting the analysis to two narrower, balanced age panels: individuals aged 50 to 70 and a stricter cohort aged 55 to 65. The estimates from these subsamples, reported in Table 6, yield results consistent with the main findings in Table 4. Specifically, the effect of the pension on labor supply remains negative and statistically significant for both agricultural and non-agricultural work. The sharp decrease in sample size explains slight fluctuations in the magnitude and precision of the coefficients, but the core findings remain unchanged, underscoring the overall robustness of the analysis.

Table 6: Results for the Panels with Restricted Age Ranges

Variables	Age between 50 and 70		Age between 55 and 65	
	(1)	(2)	(3)	(4)
	Agricultural	Non-Agricultural	Agricultural	Non-Agricultural
ATT	-0.075*** (0.026)	-0.039*** (0.013)	-0.049* (0.029)	-0.062*** (0.020)
Control Variables	✓	✓	✓	✓
Observations	8,896	8,896	5,444	5,444

Note: Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Control variables are same as before. All the results are estimated by CSDID.

5.7 Double Machine Learning with Fixed Effects

The validity of both standard and staggered DID estimators crucially depends on the parallel trends assumption, which is often difficult to justify empirically. Furthermore, recent advances in the econometrics literature have shown that staggered DID designs are susceptible to bias in the presence of treatment effect heterogeneity over time or across cohorts, as comparisons between earlier- and later-treated units can become contaminated. To provide more robust evidence for our results, we implement the double machine learning (DML) estimator proposed by Chernozhukov et al. (2018). This semi-parametric approach is less reliant on functional form assumptions and helps mitigate misspecification bias by using machine learning methods to flexibly estimate the nuisance parameters, specifically the propensity score and the outcome model. The DML framework specifies the data generating process (DGP) as:

$$Y_i = \theta_0 D_i + g_0(X_i) + \varepsilon_i, \quad (2)$$

$$D_i = m_0(X_i) + v_i, \quad (3)$$

where D_i is the treatment variable, X_i is a vector of pre-treatment covariates, θ_0 is the treatment effect parameter of interest, $g_0(\cdot)$ and $m_0(\cdot)$ are unknown functions, and ε_i and v_i are error terms with $\mathbb{E}[\varepsilon_i|X_i, D_i] = 0$ and $\mathbb{E}[v_i|X_i] = 0$. Taking conditional expectations on both sides of Equation (2) and then subtracting from it yields

$$Y_i - \mathbb{E}[Y_i|X_i] = \theta_0 (D_i - \mathbb{E}[D_i|X_i]) + \varepsilon_i \quad (4)$$

Let $\hat{g}(X_i) = \hat{\mathbb{E}}[Y_i|X_i]$ and $\hat{m}(X_i) = \hat{\mathbb{E}}[D_i|X_i]$ denote some machine learning estimators of $\mathbb{E}[Y_i|X_i]$ and $\mathbb{E}[D_i|X_i]$ respectively. They can be obtained by some kind of machine learning algorithm such as LASSO, random forest, boosting, and neural network. Then we follow [Chernozhukov et al. \(2018\)](#) and consider the regression model:

$$Y_i - \hat{\mathbb{E}}[Y_i|X_i] = \theta_0 (D_i - \hat{\mathbb{E}}[D_i|X_i]) + \eta_i, \quad (5)$$

where η_i is the error term that incorporates the estimation errors from the first-stage nuisance models. This orthogonalization step is key to debiasing the estimator. This approach allows us to consistently estimate the causal effect of the treatment while flexibly controlling for potential confounders, providing a more robust alternative to DID that is less vulnerable to its core identifying assumptions.

The above method can be readily adapted into a panel data setting like CHARLS dataset. [Clarke and Polselli \(2025\)](#) showed that with usual panel data methods like within group transformation or correlated random effects, one can apply DML after first transforming the data. Thus, we use the following model to estimate the effect of receiving NRPS on labor supply:

$$Q(Work_{it}) - \hat{g}(Q(X_{it})) \approx \theta^0 [Q(NRPS60_{it}) - \hat{m}(Q(X_{it}))] + \eta_{it} - \bar{\eta}_i,$$

where $Q(X_{it})$ represents the within transformation $X_{it} - \bar{X}_i = X_{it} - T^{-1} \sum_{t=1}^T X_{it}$, with other variables defined similarly. $Work_{it}$ denotes agricultural work or non-agricultural work, $NRPS60_{it}$ is the interaction term denoting whether the elderly received NRPS, and X_{it} denotes the control variables. Here, $\hat{g}(\cdot)$ and $\hat{m}(\cdot)$ can be regarded as the first-order approximation (i.e., linear function of X_{it}) of $g_0(\cdot)$ and $m_0(\cdot)$, which allows the use of [Clarke and Polselli \(2025\)](#)'s estimation procedure in CHARLS panel settings.

Table 7: Double Machine Learning with Fixed Effects

	Q(Agricultural)			
	(1) Lasso	(2) Random Forest	(3) Gradient Boosting	(4) Neural Network
Q(NRPS60)	-0.022** (0.011)	-0.030** (0.011)	-0.025** (0.011)	-0.026** (0.011)
Constant	-0.000 (0.002)	0.005** (0.003)	-0.000 (0.002)	-0.008** (0.002)
Control Variables	✓	✓	✓	✓
Observations	13,584	13,584	13,584	13,584
	Q(Non-Agricultural)			
	(1) Lasso	(2) Random Forest	(3) Gradient Boosting	(4) Neural Network
Q(NRPS60)	-0.005 (0.006)	0.001 (0.007)	-0.000 (0.006)	-0.001 (0.006)
Constant	-0.000 (0.002)	-0.002 (0.002)	-0.000 (0.002)	-0.002 (0.002)
Control Variables	✓	✓	✓	✓
Observations	13,584	13,584	13,584	13,584

Note: Robust standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. $Q(\cdot)$ refers to the within group transformation. Control variables are same as before, with their within group transformation additionally.

Table 7 presents the estimation results using the DML approach on the transformed data. The results are largely consistent with those in Table 5, though they differ from the estimates for non-agricultural work reported in Tables 4 and 6. Moreover, since DML controls for confounders

with unknown functional forms to enhance robustness, the estimated coefficients of *NRPS60* are generally smaller in magnitude across all regressions. In summary, the negative effect of NRPS receipt on agricultural labor supply is statistically significant and robust, while its effect on non-agricultural work varies across models and estimation methods.

A key advantage of DML is its ability to nonparametrically estimate the nuisance functions $g_0(\cdot)$ and $m_0(\cdot)$, which ensures that endogeneity and selection bias due to observables are effectively partialled out in the treatment effect estimation. As shown in [Chernozhukov et al. \(2018\)](#), consistent estimation of either $g_0(\cdot)$ and $m_0(\cdot)$ is sufficient for valid inference. This offers additional robustness in our context against bias arising from the nonlinear influence of observed covariates on labor supply decisions. It is important to note, however, that due to the use of within transformation in our panel setting, we restrict the machine learning models to linear functional classes. While this may limit the flexibility of DML, it aligns with conventional panel methods such as TWFE and CSDID, which also operate under linear assumptions. Thus, the DML estimates still provide a meaningful improvement in robustness compared to traditional approaches by more effectively mitigating bias from complex patterns in the observed confounders.

5.8 Test of Moral Hazard

Given that receiving the NRPS reduces agricultural labor supply, it is interesting to examine whether this additional leisure time leads to other behavioral changes among the elderly. To investigate potential moral hazard effects, such as increased engagement in health-risk behaviors, we analyze two outcomes: smoking and monthly drinking frequency. Figures [14](#) and [15](#) plot the ATT over the treatment timeline. Both outcomes exhibit some trending behavior; however, the ATT for smoking is not statistically significant.

A notable finding is evidence of an anticipation effect for drinking. Since eligibility for the NRPS is determined by an individual's 60th birthday (as recorded on one's ID card), people may begin to increase their drinking in the period immediately before they officially start receiving pension payments. Furthermore, this pre-treatment increase in drinking is primarily driven by males with agricultural *hukou*, while no significant effect is found for females.⁵

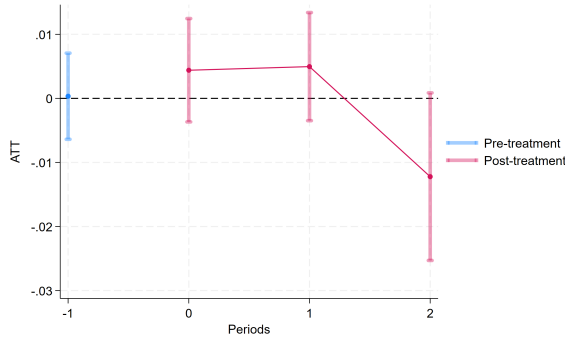


Figure 14: Event Study for Smoke

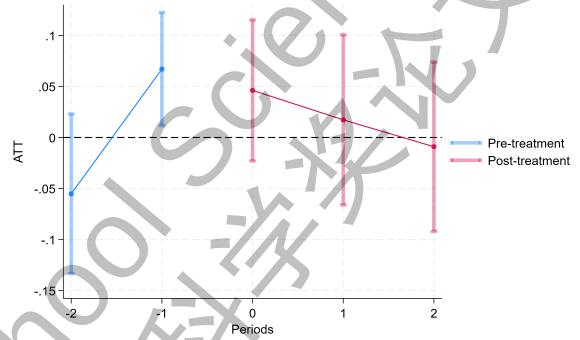


Figure 15: Event Study for Drink

Regarding the mechanism through which pension schemes affect health-related risk behaviors, [Fu et al. \(2017\)](#) identified two potential channels. First, by providing medical coverage, pensions may reduce the incentive for preventive health actions, creating an ex-ante moral hazard. Second, they may increase exposure to health information, for instance, through more frequent interactions with healthcare providers, thereby raising health awareness and encouraging preventive behaviors. Our findings align with [Fu et al. \(2017\)](#), showing that pension schemes induce relatively little change in elderly risk behaviors such as smoking and drinking.

6 Heterogeneity Analysis

Heterogeneous effects are estimated using CSDID across three categories of variables. The first dimension is gender; the results are presented in columns (1) and (2) of Table 8. Women are

⁵See Figures A1 – A4 in the Appendix.

more likely to reduce their agricultural labor supply, whereas men are more likely to reduce their supply of non-agricultural work. These findings confirm the conjecture presented in Section 4.1, which posited that women typically spend more time on agricultural activities and men on wage-earning jobs. According to the *2018 National Migrant Workers Monitoring Survey Report* issued by National Bureau of Statistics,⁶ men accounted for 65.2 percent of all migrant workers,⁷ while women accounted for 34.8 percent. Similarly, the 2015 Report⁸ reported a share of 66.4 percent for men and 33.6 percent for women. These proportions have remained largely stable over the years from 2012–2023. Also, Cai and Huang (2017), using data from the national population censuses and national agricultural censuses, documented a clear trend of feminization in China’s agricultural labour force. According to the four national population censuses conducted in 1982, 1990, 2000, and 2010, the proportion of women in agricultural employment in mainland China increased from 46.24% to 47.48%, 48.57%, and 49.22%, respectively, which is an overall rise of 2.98 percentage points over three decades.

Furthermore, columns (3) and (4) in Table 8 show that elderly individuals with agricultural *hukou* reduce their labor supply in both types of work. This result is intuitive and reinforces the robustness of the main analysis, as the NRPS is specifically designed to support those with agricultural *hukou* and without Urban Employees’ Basic Pension Insurance. Moreover, these results suggest that the differential impact of the NRPS on agricultural and non-agricultural work is primarily linked to hukou type, given that the significant negative effects are concentrated in the agricultural hukou subgroup.

⁶See https://www.stats.gov.cn/sj/zxfb/202302/t20230203_1900299.html for detail.

⁷Here migrant workers refer to individuals whose *hukou* remains rural, and are engaged either in local non-agricultural sectors or in non-local employment for six months or longer.

⁸See https://www.stats.gov.cn/sj/zxfb/202302/t20230203_1899104.html for detail.

Table 8: Heterogeneous Treatment Effects by Subgroups

	Agricultural Work					
	(1)	(2)	(3)	(4)	(5)	(6)
	Male	Female	Agricultural	Non-Agricultural	Good	Bad
			Hukou	Hukou	Health	Health
ATT	-0.007	-0.114***	-0.100***	0.114	-0.074*	-0.111***
	(0.050)	(0.032)	(0.031)	(0.099)	(0.041)	(0.040)
Control Variables	✓	✓	✓	✓	✓	✓
Observations	1,900	11,684	11,140	2,332	6,720	6,864
	Non-Agricultural Work					
	(1)	(2)	(3)	(4)	(5)	(6)
	Male	Female	Agricultural	Non-Agricultural	Good	Bad
			Hukou	Hukou	Health	Health
ATT	-0.113**	-0.023	-0.036***	-0.005	-0.053***	-0.007
	(0.046)	(0.014)	(0.013)	(0.153)	(0.018)	(0.018)
Control Variables	✓	✓	✓	✓	✓	✓
Observations	1,900	11,684	11,140	2,332	6,720	6,864

Note: Robust standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Good health refers to the health level of excellent, very good, good. Bad health refers to the health level of fair, poor, very poor. Control variables are same as before, with their within group transformation additionally. All the results are estimated by CSDID.

Finally, we conduct a sub-sample analysis based on self-reported health status. Responses of “excellent”, “very good”, and “good” are categorized as good health, while “fair”, “poor”, and “very poor” are categorized as bad health. The results, presented in columns (5) and (6), indicate that elderly individuals in poor health are more likely to reduce their agricultural labor supply. In contrast, those in good health reduce their supply in both agricultural and non-agricultural work. A potential explanation for this finding is that individuals with poor health are less likely to participate in non-agricultural work in rural areas to begin with; they are, in fact, more likely to remain at home. Consequently, the pension effect is more concentrated on their agricultural activities.⁹

Overall, the results from the heterogeneity analysis are consistent with the main findings. All statistically significant estimates are negative and are of a magnitude comparable to those from the TWFE and CSDID estimators.

7 Conclusion

This paper examines the effect of receiving benefits from the New Rural Pension Scheme (NRPS) on the agricultural and non-agricultural labor supply of the elderly in rural China. Employing a variety of methods such as the Two-Way Fixed Effects (TWFE), stacked Difference-in-Differences (CSDID) in aggregate, event-study, and calendar-time specifications, as well as Double Machine Learning (DML), we find a robust and significant negative effect of pension receipt on agricultural labor supply. The effect on non-agricultural work is also generally negative but less consistently significant across specifications. This pattern may be explained by intra-household labor division, where women (who comprise a larger share of our sample)

⁹The heterogeneous effects of NRPS receipt on drinking behavior are presented in Figures A1–A4 in the appendix and are discussed in Section 5.6.

traditionally engage more in agricultural work, while men are more involved in wage-earning jobs. Heterogeneous analysis by gender provides evidence supporting this conjecture.

The robustness of these core results is confirmed through several checks: (1) using an alternative, self-reported measure of pension receipt (OAP); (2) shortening the panel to address potential gender selection bias; and (3) restricting the age range to mitigate confounding factors. The findings remain largely consistent with the baseline model and staggered DID results.

Our results challenge the findings of [Ning et al. \(2016\)](#), who reported an increase in labor supply following pension receipt, and align with those of [Huang and Zhang \(2021\)](#), who emphasize job switching and deterrence effects. We attribute the discrepancy with [Ning et al. \(2016\)](#) to our longer panel data, which tracks respondents over more waves.¹⁰ The observed negative income effect is theoretically sound for a non-means-tested program like the NRPS, which is less susceptible to unobservable substitution effects. Furthermore, the liquidity constraint mechanism proposed by [Ning et al. \(2016\)](#) lacks empirical support in our data. Instead, our findings indicate a clear reduction in overall labor market engagement, a nuance overlooked in studies focused exclusively on job switching among employed individuals. [Huang and Zhang \(2021\)](#) suggested that, in order to finance pension contributions, rural residents may shift toward better-paid non-agricultural work, thereby leaving total labor supply unchanged. This paper, however, employs a dummy variable capturing any form of work engagement and focuses specifically on extensive margin participation. Our results provide clear evidence that receiving pension payments reduces labor market engagement among NRPS participants.

A key contribution of this paper is the application of the Double Machine Learning (DML) framework ([Chernozhukov et al., 2018](#); [Clarke and Polselli, 2025](#)) to this question, providing highly robust, semi-parametric estimates that corroborate the negative and significant effect

¹⁰[Ning et al. \(2016\)](#) have only two waves of survey data while either four or five waves of data are used in this paper.

across multiple machine learning predictors (LASSO, Random Forest, Gradient Boosting, and Neural Networks). The accumulated evidence strongly suggests that the pension payment can reduce job market participation, which likely enhances welfare among the elderly population, especially in rural areas.

Finally, we explore behavioral responses to increased leisure time and uncover an anticipation effect among men with agricultural hukou, who increase their drinking in the period immediately before receiving NRPS payments.

Policy Implications: While pension programs like the NRPS provide crucial financial security and enhance welfare by reducing old-age labor force participation, they may also inadvertently encourage harmful behaviors in the absence of guided engagement. Therefore, policymakers should consider complementing cash transfers with public health initiatives, such as increasing access to recreational facilities, promoting physical activity, and offering regular health consultations and screenings.

Appendix

The appendix presents additional tests for moral hazard by segmenting the population by gender (male and female) and hukou type (agricultural and non-agricultural). The results, shown in Figures A1 – A4, can be compared with those in Section 5.8. As previously noted, males with agricultural hukou are more likely to have increased drinking one period ahead of the NRPS receipts, while no significant effect is observed among females.

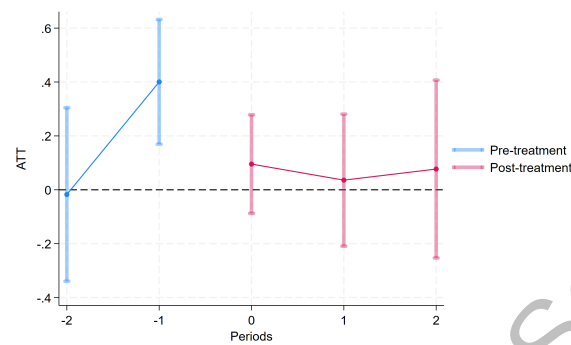


Figure A1: Event Study for Drink for Male

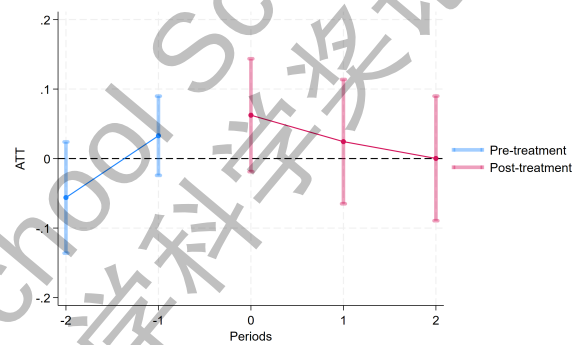


Figure A2: Event Study for Drink for Female

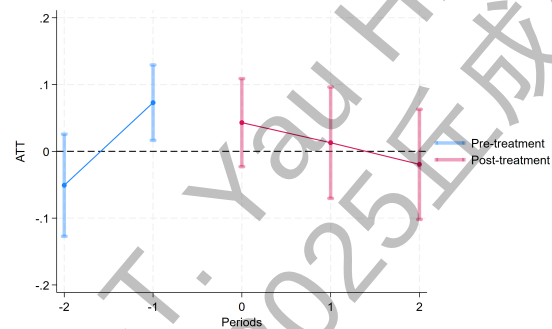


Figure A3: Event Study for Drink for Agricultural Hukou

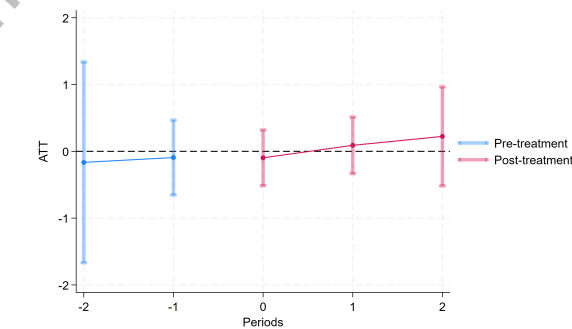


Figure A4: Event Study for Non-Agricultural Hukou

References

- Borusyak, K., Jaravel, X., and Spiess, J. (2024). Revisiting event study designs: Robust and efficient estimation. *The Quarterly Journal of Economics*, 139(2):1177–1234.
- Cai, H. and Huang, L. (2017). Who will be engaged in agriculture? investigation and reflection on sex structure of agricultural labor force. *Journal of Northwest A&F University (Social Science Edition)*, 17(02):104–112.
- Callaway, B. and Sant’ Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.
- Chang, H.-H. (2013). Old farmer pension program and farm succession: evidence from a population-based survey of farm households in taiwan. *American Journal of Agricultural Economics*, 95(4):976–991.
- Cheng, L., Liu, H., Zhang, Y., and Zhao, Z. (2018). The health implications of social pensions: Evidence from china’s new rural pension scheme. *Journal of Comparative Economics*, 46(1):53–77.
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68.
- Clarke, P. S. and Polselli, A. (2025). Double machine learning for static panel models with fixed effects. *The Econometrics Journal*, page utaf011.
- de Carvalho Filho, I. E. (2008). Old-age benefits and retirement decisions of rural elderly in brazil. *Journal of Development Economics*, 86(1):129–146.
- De Chaisemartin, C. and d’Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996.
- Fu, H., Yuan, D., and Lei, X. (2017). Health status and ex ante moral hazard of health insurance: An empirical investigation on china’s new rural cooperative medical scheme. *China Economic Quarterly*, 16(02):599–620.
- Huang, H., Zhan, J., and Chen, C. (2014). The impact of the new rural social endowment insurance pension income on the rural elderly labour supply. *Chinese Journal of Population Science*, (02):106–115.
- Huang, W. and Zhang, C. (2021). The power of social pensions: Evidence from china’s new rural pension scheme. *American Economic Journal: Applied Economics*, 13(2):179–205.
- Ma, G. and Zhou, G. (2014). The impacts of new rural pension program on household saving: Evidence from cfps. *Economic Research Journal*, 49(11):116–129.
- Meng, K. and Tuo, H. (2022). How heavy is china’s social insurance premium: Taking basic pension insurance for urban employees as an example. *Social Governance Review*, (09):53–73.

- Ministry of Finance and Ministry of Human Resources and Social Security (2011). Notice on issuing the interim measures for the financial management of the new rural social endowment insurance fund.
- Ning, M., Gong, J., Zheng, X., and Zhuang, J. (2016). Does new rural pension scheme decrease elderly labor supply? evidence from charls. *China Economic Review*, 41:315–330.
- Qian, Y. (2017). *How Reform Worked in China: The Transition from Plan to Market*. The MIT Press.
- State Council (2009). Guiding opinions of the state council on carrying out the pilot program of the new rural social endowment insurance.
- Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.
- Tang, L., Sun, S., and Yang, W. (2021). Investments in human capital: The evidence from china’ s new rural pension scheme. *Research in International Business and Finance*, 55:101345.
- Zhang, C., Giles, J., and Zhao, Y. (2015). Policy evaluation of china’s new rural pension program: Income, poverty, expenditure, subjective wellbeing and labor supply. *China Economic Quarterly*, 14(01):203–230.
- Zhao, Y., Hu, Y., Smith, J. P., Strauss, J., and Yang, G. (2014). Cohort profile: The china health and retirement longitudinal study (charls). *International Journal of Epidemiology*, 43(1):61–68.

Acknowledgements

This research project grew out of a personal curiosity about the well-being of the elderly in rural China, an interest first sparked during my Grade 10 year by reading Fei Xiaotong's seminal work, *From the Soil*. The book's insights into traditional rural society helped me understand the profound transformations occurring in contemporary Chinese villages, particularly how modern institutions like pension systems are reshaping elderly livelihoods. This academic interest was deeply personalized through my own family experience. My dad was born and raised in a poor rural village in Suqian, Jiangsu. He observed many elderly people dying of illness due to poverty before the New Rural Pension Scheme. My grandmother came to live with us after her family's land was reclaimed for industrial use. During visits with her to other villages in the region, I consistently heard how the New Rural Pension Scheme had enabled elderly residents, especially those in poor health, to significantly reduce their physically demanding farm work while maintaining financial stability, granting them a retirement that was previously unimaginable.

Motivated by these observations, I sought to determine whether this pattern held nationally rather than just in specific regions. I conducted an extensive literature review on pensions and labor supply using platforms like Google Scholar and CNKI, and taught myself the vocabulary of fixed effects and regression discontinuity from online lecture notes and Stata tutorials. The entire process, from conceptualization and literature review to data analysis, interpretation, and writing, was conducted by me.

The empirical heart of this project is the China Health and Retirement Longitudinal Study (CHARLS) dataset, which was obtained through the official Peking University website. I undertook the complex and arduous task of cleaning and harmonizing five waves of data (2011,

2013, 2015, 2018, and 2020). This process involved merging datasets by individual ID, excluding respondents who were already over the eligibility age and receiving NRPS benefits at the program's inception to better isolate its effects on the treatment group. The 2020 wave was excluded because it contained no pension or social-security information, and the pandemic and lock-down policies had sharply distorted both economic behavior and labor market conditions. The final panel consists of 3,396 individuals across 13,584 observations. The primary challenge lay in reconciling variable inconsistencies across survey waves and resolving logical errors in the raw data. Each obstacle was overcome through meticulous cross-referencing with the official CHARLS codebooks and questionnaires, ensuring the creation of a robust and reliable dataset for analysis.

I designed and executed all econometric models independently. Professor Xia Wang, whom I have known for several years from Renmin University of China, served as my academic advisor. Her role was exclusively advisory; she did not participate in data acquisition, cleaning, analysis, or writing. Her primary contribution was providing expert guidance on econometric methodology and data analysis techniques. I discussed the econometric methods I identified in the literature with her, and she would patiently offer feedback and suggest more advanced techniques, including contemporary panel data methods and machine learning tools for causal inference. Her guidance was entirely voluntary and uncompensated, and I am deeply grateful for her generosity. This paper would not have reached its current rigor without her mentorship.

Finally, I extend my gratitude to the researchers behind the CHARLS study and to the elderly individuals across China who shared their stories, both of which made this research possible.