Early Childhood Conditions and Adolescent Mental Health [*]

Bilge Erten[†] Pinar Keskin[‡] Rodrigo Pinto[§] Huihua Xie[¶] and Lianming Zhu[||]

September 20, 2022

Abstract

We investigate how early life circumstances induced by trade liberalization affect adolescent mental health in China. Our model differs from the classic Difference-in-differences (DiD) design in that it considers a moderator variable that determines the intensity with which the treatment affects the outcomes. Although the primary focus of this study is empirical, we investigate the assumptions that ensure causal interpretation to the DiD estimator commonly reported for this type of model. Our findings show that children born in prefectures more exposed to an exogenous change in international trade policy experienced a significant relative decline in the incidence of severe depression during adolescence, as classified by the CES-D scale. We show that these results are not driven by pre-existing trends in the outcomes, and that the estimated relationships are robust to controls for initial prefecture attributes and other policy changes. Improvements in parental income, early childhood investments, and maternal care provision are likely operative channels of impact.

(JEL Codes: F16, I15, J13)

[*]We thank Achyuta Adhvaryu, David Atkin, Prashant Bharadwaj, Randall P. Ellis, Martin Fiszbein, Jonathan Gruber, Brian Kovak, Jetson Leder-Luis, Dilip Mookherjee and the seminar participants at Aarhus University, Boston University, Deakin University, ILO Brown Bag Lunch Series, Society of the Economics of the Household (SEHO) Conference at UCL, University of Melbourne, Wellesley College, and Wilfrid Laurier University for their comments and suggestions. Bilge Erten gratefully acknowledges the hospitality of the Institute of Economic Development at Boston University. Any errors are our own.

[†]Northeastern University and IZA. Department of Economics, 312A Lake Hall, 360 Huntington Avenue, Boston, MA 02115. E-mail: b.erten@northeastern.edu

[‡]Wellesley College. Department of Economics, Pendleton East, 106 Central Street, Wellesley, MA 02481. E-mail: pinar.keskin@wellesley.edu

[§]UCLA, NBER and HCEO. Department of Economics, 8283 Bunche Hall, Room 8385, Los Angeles, CA 90095. E-mail: rodrig@econ.ucla.edu

[¶]Chinese University of Hong Kong, Shenzhen. School of Management and Economics, 2001 Longxiang Blvd., Longgang District, Shenzhen 518172, China. E-mail: huihuaxie@cuhk.edu.cn

[||]Osaka University. Institute of Social and Economic Research, 6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan. E-mail: lianmingzhu@iser.osaka-u.ac.jp

1 Introduction

Mental health conditions pose serious challenges in how children and adolescents learn, behave, or regulate their emotions, having a first-order impact on their human capital development. Among these conditions, depressive disorders are among the most common ones: 17 to 22 percent of adolescents experience at least one major depressive episode and 4 to 6 percent experience symptoms of severe depression (National Institute of Mental Health 2022; Centers for Disease Control and Prevention 2022).[1] Depressive disorders are estimated to be more prevalent among adolescents in low- and middle-income countries (World Health Organization 2019).[2]

Recent estimates from China report that the prevalence of depressive symptoms among adolescents is approximately 20-22 percent, which has exhibited an increasing trend since the early 2000s (Chinese Academy of Sciences 2020; Li et al. 2019).[3] Despite a growing literature documenting the effects of early life economic shocks on later life outcomes,[4] little evidence exists on the determinants of adolescent mental health, partly due to challenges in finding exogenous sources of variation in early life economic conditions. In this paper, we document a link between adolescent mental health, particularly the incidence of severe depression, and a large, plausibly exogenous shock to local labor markets in China driven by a change in international trade policy.

In January 2002, the United States effectively implemented a bill granting permanent normal trade relations (PNTR) to China, a trade liberalization that differentially exposed Chinese regions to increased access to US markets through their industry composition. Although Chinese exports to the US had already been subject prior to PNTR to low normal trade relations (NTR) tariff rates that applied to most US trade partners, access to these low tariff rates required annual renewal by the US Congress. This process

[1]According to World Health Organization (WHO) estimates, depression is projected to become the leading cause of global disease by 2030 (Mathers and Loncar 2006). Depressed adolescents are more likely to perform poorly at school, have impaired social relationships, have substance abuse problems, and experience disability and premature death (Keenan-Miller et al. 2007; Fletcher 2010; Thapar et al. 2012).

[2]The economic losses from depressive disorders are estimated to generate losses of 50 million years lived with disability, more than 80 percent of which correspond to low- and middle-income countries (World Health Organization 2017).

[3]Consequently, the National Health Commission of China released the first action plan for the prevention and control of depression among adolescents and other vulnerable groups in 2020 (National Health Commission of China 2020; Li et al. 2021).

[4]Almond et al. (2018) provide an overview of the recent literature on early-life conditions and adult health. Heckman (2012) provides a developmental approach to health focusing on the costs and benefits of interventions over the life cycle. Currie (2020) reviews empirical studies on childhood mental conditions and their long-term consequences.

entailed considerable uncertainty since if the renewal failed, Chinese exports would be subject to the much higher non-NTR rates reserved for nonmarket economies. The passage of PNTR ended the requirement for annual renewals, reducing the uncertainty associated with exporting to the US and thereby increasing the access of Chinese firms to the US market. Exploiting regional variation in tariff uncertainty faced by local labor markets before 2001 and utilizing panel data for Chinese counties from 1996 to 2013, Erten and Leight (2021) document that more exposed counties experienced an increase in exports and foreign direct investment, a decline in agricultural employment, and an increase in manufacturing and service sector employment after PNTR. These structural shifts increased local GDP, in both total and per capita terms, in comparison to that of less affected regions. Nevertheless, relatively little is known about how early life exposure to such local labor demand shocks affects psychological wellbeing later in life, including adolescence, a critical period for the onset of many mental disorders, including depression.[5]

Early life exposure to trade-induced employment and income changes may influence adolescent mental health outcomes through household income and parental time reallocation channels. First, parents in affected regions may experience an increase in their income levels, allowing them to increase their investment in children at an early age. These early life investments may include greater time spent breastfeeding, a higher number of vaccinations, and greater investments in the nutritional intake of children. Second, if male workers find better earning opportunities, their female partners may respond by working less and exiting the work force. This may increase the time that mothers spend with children when they are young, potentially improving their later life mental health observed during adolescence.

Using a nationally representative household survey from China, we examine how changes in early life circumstances induced by trade liberalization affect mental health outcomes in adolescence. We focus our attention on early life exposure to PNTR, building on the growing body of empirical work that shows large and persistent effects

[5]Studies from the psychology literature indicate that the onset of major depressive disorder (MDD) typically occurs during adolescence (Wilson et al. 2015; Kessler et al. 2005; Costello et al. 2003). While the estimates of MDD incidence in childhood range from 1-3 percent, the estimates of its incidence in adolescence increase to 4-6 percent, which are close to the levels observed in adulthood. Longitudinal studies following adolescents through adulthood such as the Oregon Adolescent Depression Project document that adolescents with MDD experience worse outcomes than unaffected youths in relationship quality, school and work functioning, and physical health, as well as greater psychiatric comorbidity and suicidality during adulthood (Rohde et al. 2013; Marmorstein et al. 2014; Hammen et al. 2008; Zisook et al. 2007; Lewinsohn et al. 2003). This evidence suggests that the benefits of early treatment of depression during adolescence are critical for the psychological wellbeing of individuals in adulthood.

of shocks, investments, and interventions in early life on health and human capital formation (Heckman 2006, 2007, 2012; Almond and Currie 2011; Almond et al. 2018).[6] In particular, we use nationally representative data from the China Family Panel Studies (CFPS), which has a module on the Center for Epidemiologic Studies Depression Scale (CES-D8), a measure of depression internationally validated for use in nonclinical settings.

We then implement a generalized difference-in-differences (DiD) identification strategy to examine whether cohorts born in prefectures more exposed to PNTR experienced differential changes in mental health outcomes during adolescence in comparison to the changes observed among cohorts born in less exposed prefectures after the policy was implemented.[7] We include controls for the initial demographic and economic characteristics of prefectures where children were born, other trade policy changes, and fixed effects that absorb time-invariant attributes of prefectures of birth and aggregate shocks affecting all prefectures in a given year of birth.

We find that cohorts born in prefectures more exposed to the reductions in trade policy uncertainty due to PNTR experienced a significant relative decline in the incidence of severe depression during adolescence, as classified by the CES-D scale. We show that these declines took place at the time of the policy change, with estimates implying that an interquantile shift in prefectures' exposure to PNTR is associated with a decline in the probability of experiencing severe depression of 3.8 percentage points, or 63 percent of the mean severe depression incidence among adolescents, in cohorts born in more PNTR-exposed prefectures relative to those born in other prefectures of China. Using a standardized mental health index, we also document a 0.1-standard-deviation improvement in the mental health of cohorts born in more PNTR-exposed prefectures in comparison to that of cohorts born in other prefectures, while we find no evidence

[6]The evidence from neuroscience literature indicates that brain development in the first years of life plays a crucial role in mental disorders given the presence of greater plasticity and neurogenesis. Total brain volume doubles in the first year of life while this increase goes down to 15 percent by the second year (Knickmeyer et al. 2008). Moreover, childhood experiences right after birth shape neural circuits in the brain that mediate socioemotional behaviors more than any other time period later in life (Knudsen et al. 2006).

[7]Prefectures are the second administrative division of China, below provinces. The most common form of the prefecture is the so-called "prefecture-level city" (*dijishi*). There are also prefectures that are not prefecture-level cities, and the term "county-level city" (*xianjishi*) is the official name for such jurisdictions. County-level cities have judicial rights but not legislative rights over their own local law and are usually governed by prefecture-level divisions. Most county-level cities were created in the 1980s and 1990s by replacing more densely populated counties. Such county-level cities are not "cities" in the strictest sense of the word, since they are usually much larger than a metropolitan area and cover rural areas many times the size of their urban, built-up area. Both metropolitan and rural areas of China are covered in this paper, and we refer to them as prefectures.

of a significant impact of the policy change on the incidence of mild depression among adolescents.

Which channels explain our findings for later life effects on mental health? We first examine the most proximate outcomes: parental income, early life investments, and nutrition intake of children. Using China Statistical Yearbook data, we document that average income per worker increases in prefectures more exposed to PNTR in comparison to less affected prefectures after the policy change. Using China Health and Nutrition Survey (CHNS) data, we find that children in more PNTR-exposed prefectures experienced a relative improvement in early life investments, such as prenatal visits, duration of breastfeeding and the number of vaccinations that they received, and in their nutrition intake in terms of total calories, protein, carbohydrate and fat consumption, which resulted in increases in child height and weight. Second, we use several rounds of the China Population Census to examine the effects of the trade policy change on employment status outcomes. Our findings show that women in more PNTR-exposed prefectures experienced a relative decline in their labor force participation.[8] Using the CHNS data, we then show evidence for a significant decline in the number of days that children are cared for by people outside the household per week and a significant decline in the hours of care that they receive from people outside the household in more exposed prefectures after the reform. These results imply that the probability of children being cared for at home, potentially by their mothers, significantly increased after the policy change in more affected prefectures. Furthermore, we show that our results are not explained by selective migration, parental absence, or fertility in response to the policy change.

Our analysis contributes to several strands of literature. First, it relates to a growing body of research on the economics of child and adolescent mental health. Recent evidence shows that mental health conditions that emerge in adolescence have long-lasting effects (Currie and Stabile 2006, 2009; Currie et al. 2010). Using administrative data from Canada, Currie et al. (2010) compare children with mental health disorders to their siblings, finding that having attention deficit hyperactivity disorder or conduct disorder at ages 14-18 substantially increases the probability of going on welfare at age 18. Indeed, Currie (2020) emphasizes the need to focus on childhood mental health and on the "missing middle" years of adolescence, on which there is far less research. Focusing on the impacts of early childhood conditions on adolescent mental health

[8]In contrast, men in more affected regions did not experience a significant change in their labor force participation while they shifted from agriculture toward manufacturing and services.

outcomes, our work contributes to this less studied topic of interest.

The link that we find between early exposure to PNTR and adolescent mental health also relates to a series of papers studying how early life conditions can affect future mental health. Most of the evidence on how individuals' early life experiences affect their later life outcomes comes from exposure to natural disasters such as disease and famine outbreaks, confirming the fetal origins hypothesis that early life access to nutrition has long-run effects on health, well-being, and economic outcomes.[9] While this evidence is crucial for evaluating the persistent effects of such early life adverse shocks, it may not compare well with the effects of more common and malleable early life experiences that may be influenced by policy changes. Our study documents later life effects observed in adolescence of an early exposure to a nationwide policy intervention.[10]

Finally, our study contributes to a growing international trade literature that documents the effects of trade policy on a range of health and economic outcomes, including mortality and marriage market outcomes (Autor et al. 2019; Pierce and Schott 2020), self-reported health assessments (Lang et al. 2019; McManus and Schaur 2016), labor market outcomes (Dix-Carneiro and Kovak 2019; McCaig and Pavcnik 2018; Pierce and Schott 2016), intimate partner violence (Erten and Keskin 2021), crime (Dell et al. 2019; Dix-Carneiro et al. 2018), and local public goods provision (Feler and Senses 2017). Our analysis adds to a broader understanding of the distributional consequences of trade liberalization by focusing on an outcome—adolescent mental health—that has not been previously studied in the trade literature. We also contribute to this literature by considering the impacts of early exposure to a trade shock during childhood on later life health and cognition.

The rest of the paper is organized as follows. Section 2 describes the data, Section 3 investigates the causality of DiD designs with moderation. Section 4 outlines our empirical strategy, and Section 5 presents our results for key outcomes and discusses

[9]Almond et al. (2018) and Currie (2020) provide an overview of this literature and how it relates to child mental health. For instance, exposure to the Dutch "Hunger Winter" during World War II or to the Six-Day War in Israel during the fetal period has been found to be associated with an increase in the likelihood of experiencing schizophrenia (Susser et al. 1998; Malaspina et al. 2008).

[10]Our study also complements previous studies in economics and public health by examining the effects of a major, plausibly exogenous shock with large local labor market effects in a fast-industrializing developing-country setting. Related studies include Persson and Rossin-Slater (2018), who find that exposure to maternal stress induced by deaths of close family members in utero leads to a large increase in the usage of prescription drugs for mental disorders during childhood and adulthood, and Adhvaryu et al. (2019), who find that exposure to favorable circumstances in early life driven by positive commodity price shocks leads to substantial declines in severe mental distress in adulthood.

mechanisms. Section 6 concludes.

2 Data

In this section, we describe the data sources employed in our analysis. In addition, where necessary, we explain how the main variables of interest are constructed.

2.1 Mental Health

Our primary measure of mental health is constructed using the 8-question Center for Epidemiologic Studies Depression Scale (CES-D8). These data were collected as part of the China Family Panel Studies (CFPS), a nationally representative biennial survey designed to complement the Panel Study of Income Dynamics in the United States. We use two waves of CFPS data collected in 2016 and 2018, which consistently asked CES-D8 questions.[11] The CES-D scale was developed by Radloff in 1977 as a validated instrument to measure depression in nonclinical settings (Radloff 1977). It has been widely used in large health surveys, such as the National Health Interview Survey and the National Household Survey on Drug Abuse, and has been validated and used in more than 30 countries, including China. The Chinese version of the CES-D scale has been widely adopted in previous research (Greenberger et al. 2000; Chen et al. 2009; Zhou et al. 2018), and its reliability and validity have been tested among Chinese adolescents (Rankin et al. 1993; Zhang and Norvilitis 2002; Chen et al. 2009). While the original CES-D scale consisted of 20 items (Radloff 1977), the 8-item version (CES-D8) is frequently used and has been shown to reliably measure depression (Kohout et al. 1993; Steffick et al. 2000). The questionnaire consists of 8 statements about several mental states experienced during the previous week. The respondents are asked to rate each item from 0 to 3, ranging from "never" to "all of the time". More specifically, respondents rate the following statements regarding how they felt during the week prior to the interview:

1. I felt depressed.
2. I felt that everything I did was an effort.

[11]The CFPS has three more rounds from 2010, 2012, and 2014. While the 2012 wave used the 20-item CES-D scale, the 2010 and 2014 rounds used the 6-item Kessler Psychological Distress Scale (K6). To have a consistent measure of mental health assessment across survey rounds and include comparable cohort sizes for treatment and control groups, we use the 8-item CES-D scale evaluated by respondents aged 10 and above in the 2016 and 2018 waves of the CFPS.

3. My sleep was restless.
4. I felt happy. (reverse coded)
5. I felt lonely.
6. I enjoyed life. (reverse coded)
7. I felt sad.
8. I could not get "going".

Using the validated cutoff points (Rushton et al. 2002; Steffick et al. 2000), we create two indicator variables to measure incidence of depression: (i) mild depression, which takes the value of one if the CES-D score ranges from 7 to 9; and (ii) severe depression, which takes the value of one if the CES-D score is greater than or equal to 10.[12]

The CES-D8 questions were answered by individuals at or above the age of 10. For our analysis, we retain individuals born in China who have nonmissing responses on the prefecture of birth and CES-D questions and who were born between 1999 and 2004, with ages ranging from 12 to 19 at the time of the survey.[13] We describe the definition of the treatment and control groups in Section 4. This leaves us with a sample of 3,551 individuals for mental health outcomes.

In addition, the CFPS includes information on physical health outcomes. We create a physical health index by taking a simple average of z scores for the following two measures: (i) an indicator variable that takes the value of one if the respondent felt physically uncomfortable during the past two weeks and (ii) an indicator variable that takes the value of one if the respondent was hospitalized last year due to illness or injury. Higher values of the index reflect better physical health.

Finally, the CFPS presents respondents with two sets of cognitive tests to evaluate their cognitive ability. While one of these tests focuses on assessing the verbal ability of respondents, the other evaluates their math ability. Using this information, we construct a cognitive function index by taking a simple average of z scores for the verbal and math test scores. Higher values of the index again reflect higher cognitive function.

[12]In addition, we create a mental health index, which is an average of the z scores of the eight reversed mental health measures included in the CES-D. We create all indices to have a mean of 0 and a standard deviation of 1, following (Anderson 2008). Higher index values reflect better mental health.

[13]We focus primarily on the period of adolescence, which is the transitional stage from childhood to adulthood and typically spans from ages 12 to 18. This period also coincides with the onset of puberty (Jaworska and MacQueen 2015).

Panel A of Table 1A provides summary statistics for the CFPS data in our sample. We observe that 15 percent of adolescents experience mild depression and 6 percent experience severe depression. The average age of adolescents in our sample is 15, and approximately 52 percent are male.

2.2 CHNS, City Statistical Yearbooks, and Census Data

We use data from the China Health and Nutrition Survey (CHNS), which was conducted by the Carolina Population Center at the University of North Carolina at Chapel Hill and the National Institute for Nutrition and Health (NINH; formerly the National Institute of Nutrition and Food Safety) at the Chinese Center for Disease Control and Prevention (CCDC). The survey uses a multi-stage, random cluster process to draw samples in 52 prefectures of 11 Chinese provinces representing wide geographic and economic variation. We use six waves in our analysis: the 2000, 2004, 2006, 2009, 2011, and 2015 waves.

The CHNS includes information on early life investments in and the nutrition intake of children. In particular, the Pregnancy History File (PHF) of the CHNS provides information on childbearing for women who were pregnant during the survey period. From this dataset, we create two measures of early life investments in children: one is the total number of months of breastfeeding that each child received, and the other is whether the respondent reported that a child received a specific vaccine. Moreover, during the child survey, the child's caregivers were asked to report the total caloric intake for each child aged 0-12 in the past 3 days and how many grams of protein, carbohydrate, and fat were consumed by each child in the past three days. We retain children who had nonmissing responses on the prefecture of residence and who were born between 1999 and 2004. This leaves us with a sample of 543 children from the PHF data and 1,650 children from the child survey. Panel B of Table 1B provides summary statistics from the CHNS data for these indicators. The average duration of breastfeeding was 10 months, and children received two vaccines on average.[14] The children, on average, consumed 1,311 $Kcal$ calories, 43 g of protein, 179 g of carbohydrates and 47 g of fat over the past three days. In addition, we also have information on height and weight for children aged 0-12. The average height for these children is 121 centimeters and the average weight is 26 kilogram.

Moreover, we use information on childcare provision and prenatal visits of pregnant

[14]Since the PHF surveyed only women who were pregnant during the CHNS sample period, the sample for early life investments is smaller than the nutrition sample.

women from the CHNS data. In particular, our childcare measurements are the number of hours per day and the number of days per week that a child is cared for by people outside the household for children aged 0 to 6 years; the summary statistics are tabulated in Panel B of Table 1A. Panel C of Table 1B presents the summary statistics for prenatal visits of mothers who had become pregnant since the previous wave of the survey.

In addition, we use data from the China City Statistical Yearbooks from 1995 to 2015 on the total labor income and the total labor income per worker in each prefecture. Panel D of Table 1B presents the summary statistics for these outcomes. This data allows us to corroborate reform-induced positive income effects documented using GDP per capita at the county level by Erten and Leight (2021).

Finally, we use data from the China population census by combining the 1990, 2000, and 2010 census waves and the 2005 and 2015 one-percent population censuses. The census contains detailed information on region of residence, employment status, industry, demographic characteristics, and educational attainment. We aggregate the individual-level data to the prefecture level and calculate the share of total employed (and its composition by agriculture, manufacturing and service sectors), unemployed, and out-of-labor-force individuals to the marriageable and working-age population for women and men separately.[15] As presented in Panel E of Table 1B, the average employment rate is 77 percent for women and 91 percent for men. Forty-six percent of women and 49 percent of men work in the agriculture sector. Approximately 10 percent of women work in manufacturing, while 20 percent of them work in service sectors. Similarly, 12 percent of men work in manufacturing, and a higher share of men—approximately 30 percent—work in service sectors. In terms of the labor force nonparticipation rate, 20 percent of women and 7 percent of men do not participate in the labor market.

2.3 Measuring Exposure to PNTR at the Prefecture Level

China's accession to the WTO was the culmination of a complex and lengthy process of negotiation. Prior to accession, China's Normal Trade Relations (NTR) status in the US market required a risky annual renewal by Congress; if the renewal failed, Chinese exports would be subject to the much higher rates reserved for nonmarket economies. For example, in 2000, the average US NTR tariff was 4 percent, but China would have faced an average non-NTR tariff of 31 percent had its status been revoked. The US granted

[15] In China, the legal marriage age is 20 for women and 22 for men, and the retirement age is 55 for women and 60 for men.

permanent NTR (PNTR) status to China as of 2002, but the status of Chinese exports in other markets did not change at that point. China's WTO membership significantly reduced uncertainty about US trade policy for China, generating a substantial increase in Chinese exports to US markets.

We utilize variation across Chinese prefectures in their concentration in different industries in 1990 and variation across industries in the gap between the lower tariffs applied to most-favored-nation tariffs and the higher nonmarket rates. On average, a prefecture covers approximately 1.4×10^4 square kilometers and had a population of 3.7 million in 2000. We use the prefecture as the geographic unit of the local labor market for the following two main reasons. First, commuting ties are strong within prefectures in China but weak across prefectures. For this reason, a prefecture in China is similar to a commuting zone (CZ), a geographic unit for defining a local labor market in the United States.[16] Another reason is that economic activities are more integrated within prefectures. The target-based performance evaluation system in China incentivizes top local bureaucrats (city mayors and Party secretaries) to implement various policies, such as investment and environmental policies, within prefecture boundaries.[17]

For each prefecture, we calculate a variable denoted "NTR gap" that is equal to the weighted average of the tariff gap across local industries operating in the prefecture; employment weights are used, constructed using each industry's share of local employment in 1990. Intuitively, a prefecture with a high NTR gap was exposed to a high level of uncertainty prior to PNTR because its key industries risked facing high tariffs, and therefore, such prefectures benefited more from the removal of uncertainty over tariffs.

$$NTR\,Gap_p = \sum_i empshare_{ip}^{1990} \times NTR\,Gap_i \tag{1}$$

where $NTR\,Gap_p$ denotes the NTR gap for prefecture p, $empshare_{ip}^{1990}$ denotes the share of employment by industry i in prefecture p in 1990, and $NTR\,Gap_i$ denotes the NTR gap for industry i, which is the difference between the higher tariff rate that would have applied in the case of revocation of China's NTR status and the lower NTR rate, $NTR\,Gap_i = Non\,NTR\,Rate_i - NTR\,Rate_i$.[18]

[16] The concept of the CZ was developed by Tolbert and Sizer (1996) and used by Autor et al. (2013).

[17] Another example of government policies implemented at the prefecture level is the household registration (*hukou*) system. Interprefecture migration is limited due to the *hukou* system in China. Less than 5 percent of the working age population changed their prefecture of residence between 2000 and 2005.

[18] We use the industry-level NTR gap data constructed by Pierce and Schott (2016) using ad valorem equivalent NTR and non-NTR rates. The NTR gap for industry i is the average NTR gap across the

Since each prefecture's sectoral composition prior to WTO accession is used to construct the employment shares, the NTR gap does not reflect endogenous changes in employment composition that are driven by reduced trade policy uncertainty. Moreover, almost all of the variation in the NTR gap is explained by variation in non-NTR rates, which were set by the Smoot–Hawley Tariff Act of 1930, implying that NTR gaps did not change in response to current economic conditions in the US or China. Prefectures characterized by a larger NTR gap experienced a greater reduction in trade policy uncertainty after 2001 and therefore were more likely to have a greater expansion in export-oriented industries. Since the PNTR rates became effective for China as of January 1, 2002, our analysis characterizes all years from 2002 onward as the post-reform period. Across prefectures, the NTR gap averages 10.6 percent and has a standard deviation of 2.2 percent, with an interquartile range from 9.4 to 10.9 percent. Figure 2 illustrates the regional variation in the NTR gap across prefectures. Darker prefectures faced the largest declines in tariff uncertainty, while lighter prefectures faced smaller declines. Overall, there is substantial variation in exposure to the reduction in tariff uncertainty across Chinese prefectures.

2.4 Other Control Variables

The CFPS and CHNS contain rich data on demographic and socioeconomic characteristics, such as gender, date of birth, marital status, and educational attainment. Panels A and B of Table 1A also provide summary statistics for children's demographic characteristics that we control for in our analysis. For the CFPS sample of children, 53 percent are boys, and the fathers of 51 percent and mothers of 39 percent of the children have completed middle school education. We observe that 54 percent of the CHNS sample children are boys; 76 percent have fathers and 67 percent have mothers who have completed middle school education.

Initial differences in prefectures' characteristics might have influenced children's outcomes and thus might contaminate our estimates. To alleviate this concern, we include interactions of year dummies with the following prefecture-level characteristics: GDP per capita, share of employment in manufacturing, and prefecture's distance to its nearest port.[19]

three-digit Chinese industry classification (CIC) tariff lines for that industry. We use the NTR gaps for 1999 following Pierce and Schott (2016) and Erten and Leight (2021). These NTR gaps are almost identical to those in 2000 or 2001; accordingly, the results are robust to the use of data from other years.

[19]Data on GDP per capita data come from the China City Statistical Yearbook 1990. Data on manufacturing employment share are calculated from the 1990 population census. To calculate the prefecture's

We further control for other ongoing policy reforms during the period of trade policy uncertainty that might have affected children's outcomes. First, we control for trade policy reforms in China, which include changes to output and input tariffs, export licenses, and barriers to investment in China. We obtain data on output tariffs at the HS-6 product level from the World Integrated Trade Solution database. We then map the HS-6 products to 3-digit Chinese industries in the census using a concordance table from the National Bureau of Statistics (NBS) of China. This allows us to calculate the simple average output tariffs at the 3-digit industry level. The industry-level input tariff is calculated as a weighted average of the industry-level output tariff, using as the weight the share of inputs in the output value from the China input–output table in 1997. With these industry-level tariffs data, we then construct prefecture exposure to output and input tariffs as the weighted average of industry-level tariff source data using the share of employment by industry in 1990.

There were severe restrictions on direct exporting by firms during the period of trade policy uncertainty. We collect data on HS 6-digit products requiring export licenses and products that could only be exported through intermediaries from the Administrative Measures Regarding Imports and Exports of China Yearbook. We then map the HS 6-digit product to the 3-digit industry level and construct a prefecture-level measure for export licenses, where higher values indicate prefectures that had higher exposure to export licenses and thus benefited more from the elimination of the restrictions on export licenses.

We proxy barriers to investment in China using an input relationship-specificity index proposed by Nunn (2007). Based on the classifications in Rauch (1999), Nunn (2007) considers goods that are neither reference priced nor sold on exchange markets to be relationship-specific goods and computes the proportion of relationship-specific inputs for each HS 10-digit product. The HS 10-digit codes are averaged to HS 6-digit products and mapped to the 3-digit industry level using the concordance table from the NBS of China. With the industry-level measures, we then construct a measure of the extent to which holdup problems affect foreign firms' production in the prefecture.

In addition, we consider another important policy change implemented during the sample period: the removal of imposed quotas on a subset of Chinese textile and clothing of textile and clothing products. We employ data on the Multifiber Arrangement (MFA) "quota-bound" product at the HS 6-digit product level from Khandelwal et al.

distance to its nearest port, we first collect information about each prefecture's latitude and longitude from China Data Online. Geographical information on port locations (specific coordinates) is extracted from the World Port Index. We can then calculate the prefecture's distance to its nearest port.

(2013). We convert from the HS 6-digit product level to the 3-digit Chinese industry level using the concordance from the NBS of China. Based on these 3-digit industry-level data, we construct a prefecture-level weighted MFA variable using employment weights from the 1990 census, where greater values represent greater exposure to quota reductions and thus greater benefits from the MFA quota removal.

Finally, we control for trade policy changes in the US, including the NTR rate itself, for which we construct an industry-weighted prefecture average NTR rate, where the composition of employment by industry in 1990 is used to construct the weights.

3 The Causality of the DiD Design with Moderation

This section investigates the causal content of DiD estimators in the presence of a moderator variable. The aim is to contribute to the literature on this topic, as the causal interpretation of such estimators is seldom addressed in most empirical papers. We refer to the next section for readers most interested in our empirical findings.

The canonical difference-in-differences (DiD) design stems from a two-period model. It compares the outcome difference before and after the intervention between a treatment and a control group. The design can identify the causal effect of the intervention on the treatment group under the assumption of common time trends. In this basic version, the DiD design is not capable of assessing additional information on how the treatment affects the outcomes. However, it is often the case that the researcher has access to a *moderator variable* that dictates the intensity that the treatment impacts the outcomes.

There are plenty of economic examples in which a moderator variable naturally arises. For instance, the effect of a rent-control policy on real-state prices depends on the share of dwellings being rented; the share of households with children moderates the effect of improving public schools on housing prices. Furthermore, the effect of a policy that confiscates firearms on crime depends on the share of the population that possesses guns. In our case, the economic effect of a trade liberalization policy on Chinese prefectures depends on how trading tariffs impact the industry of each prefecture.

The DiD designs that include a moderator variable M are commonly evaluated by the two-way fixed effect (TWFE) regression:

$$Y_{it} = \theta_t + \eta_i + \beta_{DiD} \cdot W_i \cdot Post_t + v_{it}, \text{ where } \quad W_i = M_i \cdot D_i, \tag{2}$$

where θ_t denotes time fixed effects, η_i stands for unit fixed effects, υ_i is the unobserved error term, and β_{DiD} is the DiD parameter of interest. The term $W_i \cdot Post_t$ denotes the interaction of a dummy for the post-treatment period ($Post_t$) and a variable that measures treatment intensity (W_i), which is often the multiplication of the moderator variable M_i and the treatment indicator D_i.

The literature on policy evaluation offers numerous examples of works that employ the TWFE estimator in (2). Recent examples of such settings in the literature on early childhood interventions are: Adhvaryu, Fenske, and Nyshadham (2019), who studies the impact of cocoa price variation during early life on mental well-being in adulthood; Anders, Barry, and Smithz (2021), who examine the impact of early childhood education on criminal activity; and Barr and Smith (2019), who evaluates the effect of nutritional assistance in early childhood on violent behavior in adulthood. The recent work of Khanna, Murathanoglu, Theoharides, and Yang (2022) leverages the impact of the 1997 Financial Crisis to study the labor market outcomes of migrants in the Philippines. The economic impact of the crisis is moderated by each province's average exchange rate shock.

Although the main focus of this paper is empirical, we study the theoretical aspects of the DiD design that includes a moderation variable. We term this model as moderated difference-in-differences (MDiD). Our theoretical contributions are three-fold: (1) we determine the causal parameters of interest for the MDiD; (2) we establish the assumptions that secure the identification of causal effects in the MDiD model, and (3) we investigate the interpretation and outline a few pitfalls of the TWFE estimator in (2).

Our analysis benefits from recent literature investigating the causal implication of DiD models that deviate from the canonical setup designs. Examples of these works are the fuzzy designs of de Chaisemartin and D'Haultfoeuille (2018), and staggered designs of Borusyak et al. (2022), Goodman-Bacon (2021), de Chaisemartin and D'Haultfoeuille (2020), and Sun and Abraham (2021). In particular, the MDiD model is closely related to the DiD model with a continuous treatment studied by Callaway et al. (2021). We benefit from their work extensively.

3.1 Examining a Two-period DiD Model with a Moderator

We use the canonical two-period DiD model as our leading example. We then extend the canonical setup to a more complex setting that includes the moderator variable. Our DiD setup consists of three observed variables indexed by unit $i \in \mathcal{I}$ and period $t \in \mathcal{T}$:

1. Y_{it} denotes a real-valued outcome for unit i at period t.

2. $D_i \in \{0,1\}$ is the treatment indicator for unit i.

3. $M_i \in \mathcal{M}$ is the moderator of the intensity of the treatment for unit i.

The support of the moderator variable M is given by $\mathcal{M}$ and may denote a continuous or categorical random variable. For the sake of exposition, we assume that the moderator takes positive values such that higher values are associated with a more significant impact of the treatment on the outcome. The period before the intervention is $t-1$, while the period after the intervention is t. We assume that no unit i is treated in period $t-1$. In the second period, all units i such that $D_i = 1$ are treated. We use $\Delta Y_{it} \equiv Y_{it} - Y_{it-1}$ for the outcome time-difference for unit i. The observed data $(Y_{it}, Y_{it-1}, D_i, M_i, \Delta Y_{it})$ denote the realized values of the random variables $(Y_t, Y_{t-1}, D, M, \Delta Y_t)$ for unit $i \in \mathcal{I}$. We employ the language of potential outcomes of the Holland-Rubin Causal (Holland 1986; Rubin 1978) to define two counterfactuals of interest:

1. $Y_{it}(d)$ is the potential outcome of unit i in period t when the treatment D is fixed[20] to $d \in \{0,1\}$.

2. $Y_{it}(d,m)$ is the potential outcome of unit i in period t when the treatment D is fixed to $d \in \{0,1\}$ and the moderator M is fixed to $m \in \mathcal{M}$.

We also assume the following conditions:

Assumption A.1. (No Anticipation) For all units $i \in \mathcal{I}$,

$$\text{First Period: } Y_{it-1} = Y_{it-1}(0) = Y_{it-1}(0, M_i),$$
$$\text{Second Period: } Y_{it} = Y_{it}(D_i) = Y_{it}(D_i, M_i).$$

Assumption A.2. (No Treatment Moderation) There is a set $\mathcal{M}_0 \subset \mathcal{M}$ such that for any value $m_0 \in \mathcal{M}_0$ we have that:

$$Y_{it}(d, m_0) = Y_{it}(0, m_0); d \in \{0,1\} \text{ for all units } i \in \mathcal{I}.$$

Assumption **A.1** expresses observed outcomes in terms of the outcome counterfactuals. The first equality states that the observed outcome in period $t-1$ is the untreated

[20]See Heckman and Pinto (2015) for a discussion on the fixing operator and Heckman and Pinto (2022) for a recent survey on causality.

counterfactual outcome. The second equality states that the observed outcome in period t depends on the treatment status of unit i. Note that the outcome of period $t-1$ does not depend on the treatment status D_i. Thus, the expectation of the treatment status of the second period does not affect the outcome of the first period.

Assumption **A.2** states that there exist values of the moderator that makes the treatment ineffective.[21] It is often the case that larger moderator values are associated with larger treatment effects. In our empirical setting, M moderates the impact of trade tariffs on the economy of the Chinese prefecture. The policy D changes tariffs for some industries. Its economic impact on the prefecture depends on the share of the prefecture's industries targeted by the policy. Therefore, the policy has limited effect on prefectures with closed economies or those whose industries are not affected by the policy.

We complete our basic setup stating the full support assumption:

Assumption A.3. (Full Support) $0 < P(D = 1|M = m) < 1$ for all $m \in \mathcal{M}$.

The full support assumption states that there exist treatment and control units for each value of the moderator M. This assumption will be necessary for some of our results.

The primary average effects for a binary treatment $D \in \{0, 1\}$ are:

$$ATT_t = E(Y_t(1) - Y_t(0)|D = 1) \tag{3}$$

$$ATE_t = E(Y_t(1) - Y_t(0)) \tag{4}$$

The average treatment effect on the treated, ATT_t, is the causal effect of the treatment in the outcomes of period t conditioned on the treated units. The average treatment effect, ATE_t, is the treatment effect at period t across all units. The advent of a moderator M enables us to define the following conditional effects:

$$ATT_t(m|m') = E(Y_t(1, m) - Y_t(0, m)|D = 1, M = m'), \tag{5}$$

$$ATE_t(m|m') = E(Y_t(1, m) - Y_t(0, m)|M = m'). \tag{6}$$

The parameter $ATT_t(m|m')$ denotes the treatment on the treated when we fix the moderator at the value m while conditioning on the units that share the moderator value of $M = m'$. Similarly, $ATE_t(m|m')$ is the average treatment effect when the moderator is

[21]If we set the moderator value m_0 to zero, than we have that $Y_{it}(d, 0) = Y_{it}(0, 0); d \in \{0, 1\}$ for all units $i \in \mathcal{I}$.

fixed at the value m conditioned on units i such that $M_i = m'$. We can obtain the average effects ATT_t in (3) and ATE_t in (4) by integrating the conditional effects $ATT_t(m|m)$ in (5) and $ATE_t(m|m)$ in (6) over the associated probability distribution of M. For instance, the average causal effect is given by:

$$ATE_t = \int ATE_t(m|m)dF_M(m),$$

where $F_M(m) = P(M \leq m)$ denotes the cumulative distribution function of the moderator.

We assume that the conditional effects in (5) and (6) are differentiable, which enable us to define the following marginal average effects:

$$MATT_t(m) = \left.\frac{\partial ATT_t(m'|m')}{\partial m'}\right|_{m'=m} \tag{7}$$

$$MATE_t(m) = \left.\frac{\partial ATE_t(m'|m')}{\partial m'}\right|_{m'=m} \tag{8}$$

The marginal treatment of the treated $MATT_t(m)$ in (7) measures the variation of ATT_t for a marginal change in the moderator M, while the marginal average treatment effect $MATE_t(m)$ in (8) measures the variation in ATE_t for a marginal change in M. Otherwise stated, $MATT_t(m)$, $MATE_t(m)$ measure the slope of $ATT_t(m|m)$, $ATE_t(m|m)$ with respect to M at the point $m \in \mathcal{M}$. The relationship between average and marginal effects is not as straightforward. Consider the case of a continuous moderator M whose support is given by the interval $\mathcal{M} = [\underline{m}, \overline{m}]$. Note that the integral of $MATT_t$ in (7) or $MATE_t$ in (8) over the support of the moderator M does not deliver ATT_t or ATE_t. In the case of $MATE_t$, we have that:

$$\int_{\underline{m}}^{\overline{m}} MATE_t(m)dm = ATE_t(\overline{m}, \overline{m}) - ATE_t(\underline{m}, \underline{m}). \tag{9}$$

The next proposition clarifies the relationship between the average, marginal and conditional effects:

Proposition P.1. Consider a DiD model where **A.1**–**A.2** holds, M is a continuous random variable in $[\underline{m}, \overline{m}]$, and $ATE_t(m|m)$ in (6) be a differentiable function. Then for any value $m^* \in [\underline{m}, \overline{m}]$ we have that:

$$ATE_t = \int_{\underline{m}}^{\overline{m}} MATE_t(m)\Big(\mathbf{1}[m > m^*]\,(1 - F_M(m)) - \mathbf{1}[m < m^*]F_M(m)\Big)dm + ATE_t(m^*|m^*). \quad (10)$$

Moreover, for any value $m_0 \in \mathcal{M}_0 \subset [\underline{m}, \overline{m}]$ we have that:

$$ATE_t = \int_{\underline{m}}^{\overline{m}} MATE_t(m)\Big(\mathbf{1}[m > m_0]\,(1 - F_M(m_0)) - \mathbf{1}[m < m_0]F_M(m)\Big)dm, \quad (11)$$

$$\text{and if } \underline{m} \in \mathcal{M}_0, \text{ then } ATE_t = \int_{\underline{m}}^{\overline{m}} MATE_t(m)\,(1 - F_M(m))\,dm. \quad (12)$$

If $ATT_t(m|m)$ in (5) is differentiable, then (10)–(12) also hold if we were to replace ATE_t, $MATE_t$, $ATE_t(m^*|m^*)$, $F_M(m)$ by ATT_t, $MATT_t$, $ATT_t(m^*|m^*)$, $F_{M|D=1}(m)$ respectively.

Proof. See Appendix A.1. □

Proposition **P.1** shows that is it possible to express the average treatment effect in terms of the marginal response $MATE_t$ and the conditional effect $ATE_t(m|m)$. The proposition states that for any value m^* of the moderator, the difference between ATE_t and the conditional effect $ATE_t(m^*|m^*)$ can be expressed as a weighted average of the marginal effect $MATE_t(m)$ over the moderator's probability distribution. Moreover, the weights for $MATE_t(m)$ such that $m > m^*$ are positive and given by $(1 - F_M(m))$, while the weights for $MATE_t(m)$ such that $m > m^*$ are the negative values of the CDF $F_M(m)$. If we set m^* to a value m_0 where no treatment moderation occurs, then ATE can be expressed as a function of its marginal effect as in equation (11). The weights of this equation can be further simplified into equation (12) if the lowest value of the moderator $\underline{m}$ renders the treatment ineffective.

3.2 Identification of Causal Parameters

We examine the identification of causal effects under parallel trend assumptions.

Assumption A.4. (Conditional Parallel Trends)

$$E[Y_t(0) - Y_{t-1}(0)|D = 0, M = m] = E[Y_t(0) - Y_{t-1}(0)|D = 1, M = m]\; \forall\, m \in \mathcal{M}.$$

Assumption **A.4** is a conditional version of the standard parallel trends assumption of the canonical DiD model. It states that the temporal trend of the untreated counterfactuals conditioned on the moderator M is the same for both the treatment and the control groups. The assumption renders the identification of the ATT_t effects:

Theorem T.1. Under Assumptions **A.1** and **A.4**, $ATT_t(m|m)$ in (5) and ATT_t in (3) are identified by the following equations:

$$ATT_t(m|m) = E[\Delta Y_t|D = 1, M = m] - E[\Delta Y_t|D = 0, M = m], \tag{13}$$

$$\text{and } ATT_t = \int_m ATT_t(m|m)\frac{P(D = 1|M = m)}{P(D = 1)}dF_M(m). \tag{14}$$

Proof. See Appendix A.2. □

The identification equation for $ATT_t(m|m)$ in (13) stems from the standard arguments of the canonical DiD model. The treatment-on-the-treated effect is identified by the difference between the temporal change in outcomes for the treated and the control (untreated) units.

A common goal of the empirical evaluation is to examine how the moderator affects the effect of the treatment on the outcomes. The natural procedure to assess the impact of the moderator is to compare the treatment effects across the values of the moderator M. Next proposition presents the causal interpretation of the difference of the treatment on the treated effect $ATTt(m|m)$ across moderator values:

Proposition P.2. Under Assumptions **A.1** and **A.4**, we can decompose the difference between the treated on the treated as:

$$ATT_t(m|m) - ATT_t(m'|m') = \underbrace{ATT_t(m|m) - ATT_t(m'|m)}_{\text{Effect Difference}} + \underbrace{ATT_t(m'|m) - ATT_t(m'|m')}_{\text{Selection Bias on the Moderator}}. \tag{15}$$

Proof. See Appendix A.5. □

Proposition **P.2** shows that the difference between the treated on the treated effects comprises two terms. The first term is the difference in the treatment effect when we fix the moderator at different levels for the same units i such that $M_i = m$. It accounts for the change in the treatment-on-the-treated effect due to a shift in the moderator.

The second term in (15) is due to selection bias. It accounts is the change in the treatment on the treated effect between two sets of units. The parameter $ATT_t(m'|m)$ denotes the treatment effect when we fix the mediator M to the value $m' \in \mathcal{M}$ for units i that share the moderator value m. The parameter $ATT_t(m'|m)$ also fixes the mediator M to the value $m' \in \mathcal{M}$, however this effect is evaluated for a different set of units i that share the moderator value m'.

The main conclusion of Proposition **P.2** is that the Conditional Parallel Trends Assumption **A.4** is not sufficiently strong to render a clear causal interpretation of the differences between the treatment on the treated effects. The next proposition states that Conditional Parallel Trends Assumption **A.4** is not sufficient to secure the identification of the average treatment effect either:

Proposition P.3. Assumptions **A.1** and **A.4** are not sufficient to identify ATE_t in (4) or $ATE_t(m|m)$ in (6).

Proof. See Appendix A.3. □

The following assumption secures the identification of average treatment effects.

Assumption A.5. (Strong Parallel Trends) For all $m \in \mathcal{M}$ and each $d \in \{0, 1\}$,

$$E[Y_t(d, m) - Y_{t-1}(0, m)|M = m] = E[Y_t(d, m) - Y_{t-1}(0, m)|D = d, M = m].$$

Assumption **A.5** states that conditioned on the moderator M, the difference in mean outcome over time for treated and control units is the same. Assumption **A.5** rules out a particular type of selection bias. In the case of the control units, **A.5** implies that $E[Y_t(0, m) - Y_{t-1}(0, m)|M = m, D = 0] = E[Y_t(0, m) - Y_{t-1}(0, m)|M = m]$. The equality compares the time trend of the outcome when the treatment of the second period is fixed to zero. Given the moderator $M = m$, the average time trend on the outcome for control units ($D_i = 0$) would be the same as in the treated units if they had not been treated.

In the case of the treated units, **A.5** implies that $E[Y_t(1, m) - Y_{t-1}(0, m)|M = m, D = 1] = E[Y_t(1, m) - Y_{t-1}(0, m)|M = m]$. Given the moderator $M = m$, the average time trend on the outcome for the treatment group $D_i = 1$ is the same as the trend for all units i.

Mathematically, Assumption **A.5** is not strictly stronger than Assumption **A.4**. Indeed **A.5** does not implies **A.4** or vice-versa. Despite of this fact, Assumption **A.5** imposes an empirical constraint that is far more difficult to be satisfied. The benefit of the Strong Parallel Trends Assumption **A.5** is that, under full support **A.3**, ATE_t and ATT_t are the same. Therefore we can identify ATE_t in the same fashion of **T.1**:

Theorem T.2. Under **A.1**, **A.5** and **A.3**, $ATE_t(m|m)$, and ATE_t are identified by:

$$ATE_t(m|m) = E[\Delta Y_t|D = 1, M = m] - E[\Delta Y_t|D = 0, M = m], \tag{16}$$

$$\text{and } ATE_t = \int_m ATE_t(m|m) dF_M(m). \tag{17}$$

Proof. See Appendix A.4. □

3.3 What is the Causal Content of the TWFE Estimator?

In this section, we investigate the causal content of popular DiD regressions commonly used in empirical works employing moderator variables. Some shorthand notation helps discuss this topic. We use $\overline{M}_d = E(M|D = d); d =\in \{1, 0\}$ for the expected value of the moderator condition on the treatment group. Let $\overline{\Delta Y_t}_d = E(Y_t - Y_{t-1}|D = d); d =\in \{1, 0\}$ denotes the expected value of the outcome time difference condition on the treatment groups. Finally, let $P_d = P(D = d); d \in \{0, 1\}$ denotes the probability of each treatment group.

As mentioned, the DiD model with a moderator variable is commonly estimated by the two-way fixed effects (TWFE) regression in Equation (2). In the two-period model, the TWFE estimator is numerically the same as the coefficient from an Ordinary Least Square (OLS) regression of the outcome time-difference ΔY_t on a constant term α and the interaction between the treatment indicator D and the moderator variable M, that is, $W = D \cdot M$:

$$\Delta Y_{it} = \alpha + \beta_{DiD} \cdot W_i + \epsilon_i \tag{18}$$

Standard OLS assumptions are that the observed data $(Y_{it}, Y_{it-1}, D_i, M_i)$ denote random variables that are independent and identically distributed (i.i.d.) across i and ϵ_i is an i.i.d. unobserved mean-zero exogenous error term that is statistically independent of D_i, M_i. The following theorem examines the causal content of this regression:

Theorem T.3. Under standard OLS assumptions, the expected value of the OLS estimator for moderator DiD parameter β_{DiD} in (18) is given by:

$$\beta_{DiD} = \frac{\text{Cov}(\Delta Y_t, M|D = 1) + \left(\overline{\Delta Y_t}_1 - \overline{\Delta Y_t}_0\right) \cdot \overline{M}_1 \cdot P_0}{\text{Var}(M|D = 1) + \overline{M}_1^2 \cdot P_0} \tag{19}$$

Moreover, consider replacing the moderator M by a linear transformation $M^* = M - \overline{M}_1$. Then, under Assumptions **A.1** and **A.4**, the expected value of the OLS estimator for β_{DiD} in (18) is given by:

$$\beta^*_{DiD} = \int \frac{\partial E(Y_t(1) - Y_{t-1}(0)|D = 1, M^* = m)}{\partial m} \, \omega(m) \, dm \tag{20}$$

$$\text{where } \omega(m) = \frac{E(M^*|M^* > m, D = 1)\left(1 - F_{M^*|D=1}(m)\right)}{\text{Var}(M|D = 1)}. \tag{21}$$

Proof. See Appendix A.6. □

Equation (19) is a statistical result that arises from applying the Frisch–Waugh–Lovell Theorem (1933; 1963). This result can also be understood as an ANOVA decomposition that rewrites the OLS coefficient as a weighted average of the intra- and between-groups regression coefficients (see, for instance, Section 4.2 of Yitzhaki (2013)). Finally, if $M_i = 1$ for all $i \in \mathcal{I}$, then the parameter β_{DiD} in equation (19) yields the standard DiD estimator for the TWFE model, namely, $\beta_{DiD} = \overline{\Delta Y}_{t1} - \overline{\Delta Y}_{t0}$.

Equation (20) shows that, under a linear transformation of the moderator, it is possible to express the β_{DiD} in (18) as the weighted average of the time difference of the counterfactual outcomes for the treatment group. The weights $\omega(m)$ in (21) are a function of truncated expectation and the CDF of the moderator. These weights are always positive and sum to one. However, the estimated parameter cannot be easily described in terms of the marginal effects in (7)–(8).

It is instructive to compare the DiD estimator in (18) with a more general approach that uses a weighted least squares regression of the outcome on the complete set of interactions among the post-treatment period dummy, the moderator and the treatment indicator, in addition to time and unit fixed effects:

$$Y_{it} = \theta_t + \eta_i + \gamma \cdot (D_i \cdot Post_t) + \kappa \cdot (M_i \cdot Post_t) + \beta_{DiD} \cdot (W_i \cdot Post_t) + \upsilon_{it}, \tag{22}$$

where ω_i denotes the weight associated with each unit i. In the two-period model, the DiD estimator described by equation (22) is numerically the same as the coefficient from a regression of the outcome time-difference ΔY_t on a constant term α, the treatment indicator D, the moderator M, and their interaction $W = D \cdot M$:

$$\Delta Y_{it} = \alpha + \gamma \cdot D_i + \kappa \cdot M_i + \beta_{DiD} \cdot W_i + \upsilon_i, \tag{23}$$

The following theorem examines the causal content of the regressions in (22)–(23):

Theorem T.4. Under standard OLS assumptions and uniform weights ω_i across units $i \in \mathcal{I}$, the expected value of the OLS estimator for moderator DiD parameter β_{DiD} in (22)–(23) is given by:

$$\beta_{DiD} = \frac{\text{Cov}(\Delta Y_t, M|D = 1)}{\text{Var}(M|D = 1)} - \frac{\text{Cov}(\Delta Y_t, M|D = 0)}{\text{Var}(M|D = 0)} \tag{24}$$

Consider the regression weights $\{\omega_i\}_{i \in \mathcal{I}}$ that assign the probability $P(M = m|D = 1)$ for each unit i such that $M_i = m$. Under Assumption **A.1**, Full Support **A.3**, and Parallel

Trends **A.4**, the expected value of the OLS estimator is given by:

$$\beta_{DiD} = \int MATT_t(m) \frac{E(M - E(M|D = 1)|M > m, D = 1) \left(1 - F_{M|D=1}(m)\right)}{\text{Var}(M|D = 1)} dm. \quad (25)$$

Finally, consider the regression weights $\{\omega_i\}_{i \in \mathcal{I}}$ that assign the probability $P(M = m)$ for each unit i such that $M_i = m$. Under Assumption **A.1**, Full Support **A.3**, and Strong Parallel Trends **A.5**, the expected value of the OLS estimator for β_{DiD} in (22)–(23) is given by:

$$\beta_{DiD} = \int MATE_t(m) \frac{E(M - E(M)|M > m) (1 - F_M(m))}{\text{Var}(M)} dm. \quad (26)$$

Proof. See Appendix A.7. □

Theorem **T.4** states that the weighted OLS estimator for β_{DiD} in (22)–(23) has a causal interpretation of a weighted average of the marginal response of the ATT or the ATE effects depending on the adopted weighting scheme. By setting the weights to $P(M = m|D = 1)$, we are effectively modifying the distribution of the moderator of the control group to be equal to the distribution of the treatment group. Under the parallel trend assumption, we obtain a weighted average of the $MATT_t$.

The last weighting scheme in **T.4** balances the probability distribution of the moderator between treated and control groups. It sets the conditional probability distribution of the treated $P(M|D = 1)$ and the control $P(M|D = 0)$ groups to the unconditional distribution $P(M = m)$. In practice, this can be obtained by estimating the DiD parameter using an inverse probability weighting scheme where each data entry $(\Delta Y_{it}, D_i, M_i)$ is weighted by the inverse of the probability $P(M = m|D = d)$ such that $D_i = d \in \{0, 1\}$ and $M_i = m \in (M)$. Under the strong parallel trend assumption, we obtain a weighted average of the $MATE_t$.

An empirical restriction that prevents the implementation of the estimator in (22) or (23) is that a control group is frequently unavailable in DiD models that employ moderation variables. Namely, $D_i = 1$ for all $i \in \mathcal{I}$. A well-known example is Card (1992), who studies the impact of the federal policy of minimum wage increase on labor market outcomes. The effect of the policy on teenagers' wages and employment is moderated by the fraction of workers initially earning less than the new minimum wage. The units of analysis are the US states, and since the intervention is a federal policy, all states belong to the treatment group. Under no control group, our leading regression model in (2) becomes:

$$Y_{it} = \theta_t + \eta_i + \beta_{DiD} \cdot M_i \cdot Post_t + v_{it}. \tag{27}$$

The OLS estimator of β_{DiD} in (27) is obtained by regressing the outcome time-difference on the moderator M, that is:

$$\Delta Y_{it} = \alpha + \beta_{DiD} \cdot M_i + \epsilon_i \tag{28}$$

It is possible to obtain a causal interpretation of the linear regression above by invoking strong functional form assumptions. For instance, suppose that a simple linear model determines the counterfactual outcomes without treatment heterogeneity or selection bias:

$$Y_{it}(d) = \kappa_i + \tau_t + \beta_d \cdot M_i + v_{it} \text{ for } d \in \{0, 1\} \text{ and all any } t, \tag{29}$$

where v_{it} denotes a mean zero exogenous error term that is statistically independent of the moderator M and the time periods. In this case, the causal effect of the treatment on the outcome for each agent i is given by $Y_{it}(1) - Y_{it}(0) = (\beta_1 - \beta_0) \cdot M_i$. The average effect for a given value $m \in \mathcal{M}$ of the moderator is $ATE_t(m|m) = ATT_t(m|m) = (\beta_1 - \beta_0) \cdot m$. Moreover, we can express the outcome time difference as:

$$\Delta Y_{it} = (\tau_t - \tau_{t-1}) + (\beta_1 - \beta_0) \cdot M_i + (v_{it} - v_{it-1}). \tag{30}$$

In this simple model, the OLS estimator for β_{DiD} in (28) evaluates the parameter $\beta_1 - \beta_0$. If we standardize the moderator M to have a mean zero and a standard deviation of one, then OLS estimator evaluates the change in the treatment effect for an increase in one standard deviation of the moderator M.

The simple model in (29) imposes stringent functional form restrictions. We can obtain a more flexible model by exploring the assumptions of No Anticipation **A.1**, No Treatment Moderation **A.2**, and Parallel Trends **A.4**. Recall that the support of the moderator M can be partitioned into $\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_0$ where the set $\mathcal{M}_0$ contains the moderator values for which the no treatment moderation assumption **A.2** holds and the set $\mathcal{M}_1 = \mathcal{M} \setminus \mathcal{M}_0$ comprises the remaining values. It is useful to define the binary variable $B_i = \mathbf{1}[M_i \in \mathcal{M}_1]$, which indicates if the moderator value of unit i belongs to set $\mathcal{M}_1$. Under this notation, we can state the following result:

Theorem T.5. Let the moderator M be standardized to have a mean zero and a standard deviation of one. If Assumptions **A.1** and **A.2**, hold, and the counterfactual outcome for the control units is given by

$$Y_{it}(0) = \kappa_i + \tau_t + \beta_t^0 \cdot M_i + v_{it}, \tag{31}$$

where v_{it} denotes a mean-zero exogenous error term that is statistically independent of D, B and M. Then the expected value of the OLS estimator for β_{DiD} in (27)–(28), evaluates a weighted average of the following parameters:

$$\beta_{DiD} = \beta_{MATT} \cdot \omega_1 + \beta_{tt} \cdot \omega_2 + \beta_\Delta \cdot (1 - \omega_2), \text{ such that } 0 < \omega_1 < \omega_2 < 1, \tag{32}$$

where the parameters $\beta_{MATT}, \beta_{tt}, \beta_\Delta$ and weights ω_1, ω_2 are given by:

$$\beta_{MATT} = \int \frac{\partial E(Y_t(1) - Y_t(0)|B = 1, M = m)}{\partial m} \omega(m) dm, \tag{33}$$

$$\text{where } \omega(m) = \frac{E(M - E(M|B = 1)|M > m, B = 1)\left(1 - F_{M|B=1}(m)\right)}{\text{Var}(M|B = 1)}$$

$$\beta_{tt} = \frac{\partial E(Y_t(0) - Y_{t-1}(0)|M = m)}{\partial m} = \beta_t^0 - \beta_{t-1}^0, \tag{34}$$

$$\beta_\Delta = \frac{E(\Delta Y_t|B = 1) - E(\Delta Y_t|B = 0)}{(} E(M|B = 1) - E(M|B = 0), \tag{35}$$

$$\omega_1 = \text{Var}(M|B = 1)P(B = 1), \tag{36}$$

$$\omega_2 = \text{Var}(M|B = 1)P(B = 1) + \text{Var}(M|B = 0)P(B = 0). \tag{37}$$

Moreover, for $P(B = 1)$ and no time trend, $\beta_t^0 = \beta_{t-1}^0$, we have that:

$$\beta_{DiD} = \int MATT_t(m) \frac{E(M - E(M|D = 1)|M > m, D = 1)\left(1 - F_{M|D=1}(m)\right)}{\text{Var}(M|D = 1)} dm. \tag{38}$$

Proof. See Appendix A.8. □

Theorem **T.5** relaxes the linearity of the counterfactual outcomes for the treated units. However, it maintains the linearity of the control units (31), which subsumes the common trend assumption in **A.4**. The theorem states that the estimator of β_{DiD} in (28) consists of a weighted average of three parameters. The first parameter, β_{MATT} in (33), is the weighted average of the marginal ATT that replaces the treatment indicator D by the binary variable B. The second parameter, β_{tt} in (34), measures the outcome time trend associated with a unit change in the moderator. Specifically, the linearity of the control implies that $E(Y_t(0) - Y_{t-1}(0)|M = m) = \beta_{tt} \cdot m$. The last parameter, β_Δ in (35), is a ratio of two DiD estimators. The numerator is the difference of the outcome time differences between the units whose moderator values lie in $\mathcal{M}_1$ versus $\mathcal{M}_0$. The denominator is the difference in differences estimator for the moderator itself. If M were a binary variable such that $\mathcal{M}_0 = \{0\}$ and $\mathcal{M}_1 = \{1\}$, then β_Δ would become the standard TWFE estimator for the canonical DiD design.

The weights ω_1 in (36) and ω_2 in (37) refer to the components of the variance of the moderator variable. The theorem assumes that the moderator M is standardized to have a mean zero and a standard deviation of one. This assumption is not binding but substantially facilitates the notation. For instance, the variance of a mean zero moderator is given by:

$$\begin{aligned}\text{Var}(M) = {} & \text{Var}(M|B=1)P(B=1) + \text{Var}(M|B=0)P(B=0) \\ & + E(M|B=1)^2 P(B=1) + E(M|B=0)^2 P(B=0)^2. \end{aligned} \tag{39}$$

Equation (39) decomposes the variance of the moderator into four positive terms. The first term is the weigh ω_1. The sum of the first and second terms is the weigh ω_2. Setting the variance to one implies that $0 < \omega_1 < \omega_2 < 1$.

Equation (38) of **T.5** considers the case of where there no moderation values that render the treatment ineffective, that is, $\mathcal{M}_0 = \varnothing$ and $P(B = 1)$. If there is no time trends, $\beta_t^0 = \beta_{t-1}^0$, then the DiD estimator evaluates a weighted average of the marginal ATT. The DiD estimator evaluates the same parameter as in (25), which refers to the regression (22) that includes the full set of interactions.

Our empirical analysis follows the empirical literature that commonly adopts the linear regression (27) to evaluate DiD designs in the presence of a moderator variable. Theorem **T.5** clarifies that assigning a causal interpretation to this DiD estimator is possible. However, the causal content of the estimator is not a simple extrapolation of the canonical DiD estimator, which uses a binary treatment and two time periods. This result corroborates the recent literature on DiD designs, which attests to the complexity of ascribing causal meaning to DiD designs that depart from the canonical case.

4 Empirical Strategy

We use a difference-in-differences specification to examine whether adolescents born in prefectures with higher NTR gaps experienced differential changes in their mental health outcomes. More specifically, we test the effects of exposure to PNTR in the year of birth on later life mental health by estimating the following specification with ordinary least squares (OLS):

$$Y_{ipbt} = \beta Post_b \times NTR\,Gap_p + \boldsymbol{Z}'_{pb}\lambda + \delta_b + \delta_p + \delta_t + \boldsymbol{X}'_{ipbt}\theta + \epsilon_{ipbt} \tag{40}$$

where subscript i denotes the individual, p the prefecture of birth, b the year of birth, and t the survey year. The dependent variable Y_{ipbt} is the outcome of individual i born in prefecture p in year b and in survey year t. This variable can be a mental health outcome, such as the incidence of severe depression, or other individual outcomes observed during adolescence. The key variable is the interaction of the prefecture-of-birth NTR gap, $NTR\,Gap_p$, and a post-PNTR dummy, $Post_b$, equal to one for those born in 2002 and after. The coefficient, β, is of primary interest because it presents the impact of exposure to PNTR in the year of birth on adolescent mental health outcomes.

The specification also controls for prefecture-of-birth initial characteristics interacted with birth year fixed effects and other trade policies interacted with birth year fixed effects denoted $\boldsymbol{Z}'_{pb}$. The prefecture-of-birth initial characteristics include log GDP per capita, share of employment in manufacturing, and log distance to the nearest port observed in 1990. The other trade policies include China's output tariffs, input tariffs, export licensing, contract intensity, and MFA quotas, and the NTR rates observed in 2001. We also include birth year fixed effects, birth prefecture fixed effects, and survey year fixed effects, which net out characteristics of birth prefectures that are time-invariant and aggregate shocks that affected all prefectures of birth in a given birth year or in a particular survey year. Finally, the specification includes a number of individual-level controls denoted $\boldsymbol{X}'_{ipbt}$. These include age fixed effects, gender and indicator variables for whether the mother completed middle school and whether the father completed middle school. We cluster standard errors at the prefecture-of-birth level to account for serial correlation in outcomes within birth prefectures. Our preferred specification includes the entire set of these control variables; however, we also show that the results are robust to adding them gradually.

In the CFPS dataset, we have information on the mental health outcomes of respondents aged 10 and older. Previous studies in the psychology literature have found that the onset of depression episodes most commonly occurs in middle to late adolescence or among young teens (12 to 14 years) and teenagers (15 to 19 years) (Hankin 2015; Avenevoli et al. 2015; Hankin et al. 1998).[22] For this reason, we restrict our analysis to adolescents aged 12 and above in the 2016 and 2018 rounds of the CFPS in our baseline

[22]Studies from the psychology literature also document that many depressive episodes experienced during adulthood represent recurrences of adolescent-onset depression (Hankin 2015; Rutter et al. 2006).

analysis. This sample restriction amounts to including adolescents born between 1999 and 2004, yielding a treatment group of three cohorts born in 2002–2004 and a control group of three cohorts born in 1999–2001. Hence, we compare three cohorts born before the policy change to three cohorts born after the policy change in our baseline results.[23]

Additionally, we estimate an event study specification to test whether pre-existing trends in the outcomes of interest drive our results. In particular, we implement the following specification to examine whether individuals born in prefectures with higher NTR gaps experienced differential changes in their outcomes after the change in international trade policy versus before:

$$Y_{ipbt} = \sum_{b=1999}^{2004} \beta_b 1\{b = t\} \times NTR\,Gap_p + \boldsymbol{Z}'_{pb}\lambda + \delta_b + \delta_p + \delta_t + \boldsymbol{X}'_{ipbt}\theta + \epsilon_{ipbt} \tag{41}$$

where the first terms on the right-hand side are interaction terms of interest, including interactions of a full set of year-of-birth dummies (excluding 2001) with the (time-invariant) prefecture-level NTR gap. All control variables are the same as in Eq. 40. This specification allows us to test whether there is a relationship between mental health outcomes and the NTR gap and to examine when any such relationship first becomes detectable. The estimates from this specification provide a test of our identifying assumption that in the absence of the PNTR policy, cohorts born in more exposed prefectures do not exhibit differential trends in their outcomes compared to those born in less affected prefectures.

5 Results

5.1 Mental Health Outcomes

We report our primary estimates of Eq. 40 in Table 2. The coefficient estimates in Panel A show no evidence of a statistically significant relationship between PNTR and the incidence of mild depression. In Panel B, the coefficient estimates are negative and statistically significant across all specifications, indicating that relative those born in less exposed prefectures, adolescents born in prefectures more exposed to the policy change experienced declines in their incidence of severe depression. As discussed in

[23] As a robustness check, we also expand our sample by adding two younger cohorts born in 2005–2006 (aged 11-10 in 2016 and 13-12 in 2018) and two older cohorts born in 1997–1998 (aged 19-18 in 2016 and 21-20 in 2018) and show that the results are consistent with our baseline estimates.

Section 4, the most rigorous specifications including individual controls and year-of-birth fixed effects interacted with prefecture initial characteristics and with other trade policies, reported in column 3, are the preferred specifications.

The coefficient estimates presented in Appendix Table A1 also indicate that adolescents born in prefectures more exposed to the policy change experienced a relative improvement in their mental health. The magnitudes of the effects on the mental health index are moderate, while the impacts on severe depression are large. The column 3 estimate in Appendix Table A1 reveals that an interquartile shift in exposure to PNTR is associated with a relative increase in the mental health of adolescents by 0.1 standard deviation (0.067×1.5). For severe depression, an interquartile shift in exposure to PNTR leads to a 3.8-percentage-point decline in the probability of experiencing severe depression, which is 63 percent of the mean.

We also test whether the estimated effects for severe depression appeared only after the policy change. The event study estimates presented in Figure 1 show that for the period prior to the policy change, the coefficient estimates on the interaction of year of birth with the NTR gap are indistinguishable from zero. This absence of differential pre-existing trends between adolescents born in prefectures more exposed to PNTR and those born in less affected prefectures offers support for our DiD strategy. In contrast, for the period after the implementation of PNTR in 2002, the coefficient estimates shift down and become significantly different from zero at the 5-percent level.

We next show that our estimates are robust both to using different definitions of the NTR gap in Appendix Table A3 and to estimating alternative regression specifications in Appendix Table A4. More specifically, the results in Appendix Table A3 show that the estimates for depression outcomes are robust to reconstructing the NTR gap by excluding industries with the highest (Panel A) or lowest value on the NTR gap (Panel B), to winsorizing the NTR gap at the 5^{th} and 95^{th} percentiles (Panel C), and to reconstructing the NTR gap by excluding nontradable industries and using only the share of tradable industries in calculating the NTR gap (Panel D). Further robustness checks in Appendix Table A4 indicate that the estimates for depression outcomes are robust to weighting the regression by the 1990 prefecture population, to controlling for prefecture-specific birth year linear trends, and to expanding the sample to include cohorts born between 1997 and 2006.

Finally, we examine whether the policy change led to any significant heterogeneous treatment effects by gender, parental education, parental absence during early childhood, and initial share of rural population at the prefecture level in 1990. The results reported in Appendix Table A5 indicate no evidence of heterogeneous treatment effects on these dimensions, including whether the adolescent is female (Panel A), whether the mother completed middle school (Panel B), whether the father completed middle school (Panel C), whether the parents were absent for at least one week when the adolescent was between 0 and 3 years old (Panel D), or whether the initial share of rural population is above the median, representing more rural locations (Panel E).

5.2 Mechanisms

In this section, we examine potential mechanisms that could explain how early exposure to the policy change may have reduced the risk of severe depression in adolescence. We divide our analysis into three subsections by focusing on the effects of the trade policy reform on the following outcomes: (a) parental income, early life investments, and nutrition intake, (b) employment and childcare provision, and (c) migration and fertility.

5.2.1 Parental Income, Early Life Investments, and Nutrition Intake

One potential mechanism through which early exposure to PNTR might have led to decreased depression in adolescence is via an improvement in parental income. Children born into households in PNTR-exposed prefectures after the policy change were likely to have more resources due to both the higher income of parents producing tradable goods and the positive local labor demand effects stimulating the income of households in nontradable sectors. Such positive income effects during gestation and infancy could have large developmental effects that persist over time, leading to better mental health outcomes in adolescence.

Using data from the China City Statistical Yearbooks, we test whether there was an improvement in average labor income per worker in prefectures that were differentially affected by the trade reform after it was implemented.[24] We estimate the following specification:

$$Y_{pt} = \beta Post_t \times NTR\,Gap_p + \delta_p + \delta_t + \mathbf{Z}'_{pt}\lambda + v_{pt} \tag{42}$$

where subscript p denotes the prefecture and t the year. The dependent variable Y_{pt} is log labor income, or log labor income per worker, in a given prefecture p in time

[24] In the absence of a reliable and consistent measure of parental income at the individual level from survey data, we use prefecture-level information from the China City Statistical Yearbooks on labor income and number of workers employed.

period t. The key variable is the interaction of the prefecture NTR gap, $NTR\,Gap_p$, and a post-PNTR dummy, $Post_t$, equal to one for the period after 2001. The terms δ_p and δ_t are prefecture and year fixed effects, respectively. Additional controls at the prefecture–year level $\boldsymbol{Z}'_{pt}$ include year fixed effects interacted with other trade policies and initial prefecture characteristics as described in Section 4. And standard errors are clustered at the prefecture level.

Table 3 reports the results. The coefficient estimates are positive and statistically significant, indicating that total labor income and labor income per worker exhibit an increase in prefectures more exposed to PNTR in comparison to less affected regions after the policy change. These results are consistent with the finding in Erten and Leight (2021) that counties with greater exposure to PNTR exhibited higher GDP per capita after the change in China's PNTR status.

Positive income shocks induced by a policy change may, in turn, result in greater early life investments in the child, such as the frequency of prenatal visits, the duration of breastfeeding and the number of vaccinations, by shifting the household's intertemporal budget and creating incentives for parents to reinforce infants' endowments (Heckman 2007). Moreover, increases in parental income may improve the nutrition intake of children by relaxing parents' budget constraints. For example, they may allow parents to purchase relatively more expensive food items such as meat and other sources of protein. These improvements in early life conditions may also positively impact the mental health outcomes of adolescents.

To estimate the impact of contemporary exposure to the trade reform on the number of prenatal visits, we estimate the following specification using data from the 1991–2015 CHNS:

$$Y_{ipt} = \beta Post_t \times NTR\,Gap_p + \delta_p + \delta_t + \boldsymbol{Z}'_{pt}\lambda + \boldsymbol{X}'_{ipt}\theta + v_{ipt} \tag{43}$$

where subscript i denotes the woman, p the prefecture of residence and t the survey year. The dependent variable Y_{ipt} is a health outcome for woman i living in prefecture p in year t. The policy exposure is captured by the interaction of the NTR gap, $NTR\,Gap_p$, and a post-PNTR dummy, $Post_t$, equal to one for the period after 2001. The terms δ_p and δ_t are prefecture and year fixed effects, respectively. The set of individual controls $\boldsymbol{X}'_{ipt}$ includes the woman's age fixed effects and middle school completion status. Additional controls at the prefecture–year level $\boldsymbol{Z}'_{pt}$ include year fixed effects interacted with other trade policies and prefecture initial characteristics as described in Section 4. We cluster standard errors at the prefecture level.

Table 4 reports the results for changes in early life investments in response to the trade policy change. The estimate in column 1 estimate is positive and significant, indicating a relative increase in prenatal visits of pregnant mothers in response to the policy change in more affected regions. The estimates in columns 2 and 3 show that infants born after the policy change in more affected prefectures are breastfed longer and receive more vaccines than infants in other prefectures.[25] The estimates for individual vaccines reported in columns 4–9 indicate that the probability of receiving all vaccines significantly increases, with the exception of polio, for which the estimate is positive but imprecisely estimated. We interpret these results as evidence of prenatal care and early life investments as a channel for the estimated impacts of early life income shocks on adolescent mental health.

In Table 8, we present estimates for whether the policy change affected the nutrition intake and development of children. The estimates in Panels A-D indicate that children in more exposed prefectures experienced an increase in their total caloric intake in the form of protein, carbohydrate, and fat in comparison to that of children living in less affected prefectures after the reform. In Panels E and D, we also observe a positive impact of the reform on average height and weight of children born in more affected regions compared to less affected ones. We interpret these results as evidence of child nutrition and development as a channel for the estimated impacts of early life income shocks on adolescent mental health.

We also explore whether the policy change affected physical health, cognitive function, or school dropout rates during adolescence.[26] The coefficient estimates in Appendix Table A2 show no evidence of a significant change in these outcomes in response to early exposure to PNTR after the reform. These findings highlight that even though improvements in childhood nutrition do not have lasting effects on the physical health and cognition of adolescents, they may nevertheless have lasting effects in reducing the risk of severe depression among adolescents.

5.2.2 Employment and Childcare Provision

Next, we focus our attention on a related set of outcomes using the census and CHNS data. Another potential mechanism could be that exposure to PNTR may have differentially affected the employment statuses of men and women, resulting in different time allocations between work outside and inside the home. If the amount of time

[25]We estimate Eq. 40 with the only difference being that we use the prefecture of residence for regional variation since the CHNS does not include prefecture-of-birth information.

[26]While dropping out of school is an extreme outcome to proxy school performance, we do not observe any other appropriate performance indicators in the CFPS dataset.

that parents spend with children increases, these early investments could bring about lasting changes in children's mental health (Chang et al. 2019; Milkie et al. 2015).

Using multiple rounds of census data from 1990 to 2015, we estimate Eq. 42 to test whether the reform affected labor market outcomes by using the share of the working-age population in a particular employment category as the dependent variable. Each estimation is weighted by 1990 prefecture population and standard errors are clustered at the prefecture level. The results in Panel A of Table 6 show that women in prefectures more exposed to PNTR experienced a relative decline in their probability of employment and a relative increase in nonparticipation (or in not being in the labor force [NILF]). These changes are driven by a significant decline in the agricultural employment of women. While some women found new employment opportunities in manufacturing and services, these increases did not offset the decline in agricultural work. In contrast, the estimates in Panel B show that men in more exposed prefectures experienced a relative increase in their probability of employment and a relative decline in nonparticipation as a result of increasing employment opportunities in manufacturing and services.

One potential consequence of women's lower participation in the labor market might be that they could spend more time with children, becoming primary caregivers. We test whether the reform affected childcare arrangements in Table 7 using CHNS data. The estimates in Panel A show that compared to children in less exposed prefectures, children born in prefectures more affected by the policy change experienced a significant decline in the hours of care per day that they received from people outside of the household on a typical day. Similarly, the estimates in Panel B indicate that children in more exposed regions experienced a relative decline in days of care per week received from people outside of the household in a typical week. Our results on increased duration of breastfeeding by mothers reported in Table 4 are consistent with increased childcare provision at home. We interpret these results together to indicate greater time spent on care provision at home as a channel for the estimated impacts of early life income shocks on adolescent mental health. These results are also consistent with previous research from the US context showing that increases in maternal time driven by reductions in women's labor market opportunities are associated with improved child health outcomes, including reductions in emotional difficulties (Page et al. 2019).

5.2.3 Migration and Fertility

Finally, we test whether the policy change affected migration and fertility patterns in more affected regions, resulting in a potentially selected sample. If the increase in export-oriented jobs attracted higher-quality workers in more exposed regions, the children of these parents might be positively selected. To test for this possibility, we use the CFPS data focusing on the adolescents for whom we have mental health outcomes and construct three indicator variables: (i) an indicator variable equal to 1 if the respondent migrated to a different prefecture from the prefecture of birth, (ii) an indicator variable equal to 1 if the respondent migrated to a different prefecture from the prefecture of residence at the age of 3, and (iii) an indicator variable equal to 1 if the respondent changed his or her residence permit from rural to urban to migrate from a rural to an urban area.

Appendix Table A6 reports the results from testing whether the PNTR had a significant impact on the migration outcomes of adolescents in our sample. The coefficient estimates are null, implying no evidence of a significant impact of early exposure to PNTR on the migration outcomes of adolescents.

An additional possibility is that the policy change may have led to parental absence in more affected regions if some parents migrated to other regions for better employment opportunities. Indeed, the restrictive household registration (*hukou*) system renders it more difficult for migrant parents to bring their children with them, creating a large pool of "left behind" children in rural areas (Heckman and Yi 2012; Tong et al. 2019). We test whether this particular trade policy change may have contributed to parental absence in more affected regions by using three indicators: (i) an indicator variable that takes the value of 1 if parents of adolescents were absent for at least one week when the adolescents were between 0 and 3 years old based on the CFPS data, (ii) an indicator variable that takes the value of 1 if the mother was not living in the household at the time of interview and was seeking employment elsewhere based on CHNS data, and (iii) an indicator variable that takes the value of 1 if the father was not living in the household at the time of interview and was seeking employment elsewhere based on the CHNS data. The coefficient estimates presented in Appendix Table A7 indicate no evidence of a significant impact of the policy change on these parental absence outcomes.

We next test whether PNTR had a significant impact on fertility outcomes. If parents in more exposed regions had higher income levels due to PNTR, they might have reduced their desired number of children, allowing them to invest more per child.

Alternatively, having more income might have allowed parents to pay the fine for having a second child, allowing them to have more children. Table A8 provides estimates for two outcomes observed in multiple rounds of census data from 1990 to 2015 at the prefecture level: (i) the number of births during the past 12 months per 1,000 women and (ii) the number of children for women living in a prefecture. We find no evidence of a significant impact of PNTR on these fertility outcomes. In sum, we conclude that our results are not explained by selective migration or fertility in response to the trade policy change.

6 Conclusion

In this study, we estimate the effects of changes in early life conditions induced by trade liberalization on adolescent mental health. Using a nationally representative sample of households in China, we find that adolescents born in prefectures more exposed to a plausibly exogenous change in international trade policy experienced a significant and economically meaningful decline in the incidence of severe depression relative to the incidence among the same birth cohort born in other regions. An interquartile shift in early exposure to the policy change reduces the relative likelihood of severe depression by 3.8 percentage points, which is slightly more than half the mean. These cohorts also experienced an improvement in overall mental health.

Exploring potential channels, we document that prefectures more exposed to the trade reform exhibit relative increases in average income per worker. In more exposed prefectures, we also find evidence of an increase in prenatal visits of mothers relative to less affected regions after the policy change. Children born in more exposed prefectures experienced an improvement in early life investments and nutritional intake in comparison to their counterparts in less affected regions after the policy change. Additionally, our findings reveal that women, on average, responded to the policy change by exiting the labor force. The increased nonparticipation of women coincides with a decline in the probability of children receiving childcare outside of home. These results suggest that the probability of children being taken care of at home, potentially by their mothers, increased after the policy change in more affected prefectures. These findings also draw attention to the importance of designing parental leave policies that provide sufficient time for parents to spend with newborns.

Recent evidence from adolescent mental health studies indicates that mental health treatments through cognitive behavioral therapy (CBT) or a combination of CBT and antidepressant medication can reduce depressive symptoms by 43 to 70 percent after 12 weeks (Lewandowski et al. 2013; Kennard et al. 2009; March et al. 2004). We show that improving access to advanced country markets by one interquartile range at the local labor market level can reduce the prevalence of severe depression by approximately 63 percent. Since the recent World Health Organization (WHO) estimates for the cost effectiveness ratio of mental health treatments for adolescents in upper middle-income settings range from US$1,000–$5,000 per healthy life year gained (World Health Organization 2021),[27] trade policies that improve employment and earnings opportunities for households as a means for preventing depression later in life might be more cost effective. The case for such policies is even stronger when the total economic costs of depression in China, which are estimated to reach US$6,264 million annually (Hu et al. 2007) (84% of which are productivity losses due to illness), are taken into account.

Overall, these results contribute to the growing body of research on adolescent mental health, documenting the importance of positive shocks in terms of income and parental time reallocation in improving later life outcomes. This is particularly important in developing countries in which these dimensions of shocks have received less attention and resource constraints for improving mental health are more binding. Our findings also highlight the importance of potential mental health improvements stemming from early exposure to trade reforms that enhance access to advanced country markets. Previous estimates for the welfare importance of these reforms, while already large, are underestimated to the extent that they do not account for mental health.

[27]This cost effectiveness ratio is based on the WHO recommendation of implementing universal, school-based social-emotional learning programs to improve mental health and prevent suicide in adolescents. The targeted interventions at the school level for such programs are estimated to have a higher cost effectiveness ratio of US$10,000–$50,000 per health life year gained. Note also that China is an upper middle-income country by the WHO's definition.

References

Adhvaryu, A., J. Fenske, and A. Nyshadham (2019): "Early life circumstance and adult mental health," *Journal of Political Economy*, 127, 1516–1549.

Almond, D. and J. Currie (2011): "Human capital development before age five," in *Handbook of Labor Economics*, Elsevier, vol. 4, 1315–1486.

Almond, D., J. Currie, and V. Duque (2018): "Childhood circumstances and adult outcomes: Act II," *Journal of Economic Literature*, 56, 1360–1446.

Anders, J., A. Barry, and A. A. Smithz (2021): "The Effect of Early Childhood Education on Adult Criminality: Evidence from the 1960s through 1990s," *Working Paper*.

Anderson, M. L. (2008): "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 103, 1481–1495.

Autor, D., D. Dorn, and G. Hanson (2013): "The China syndrome: The impact of import competition on U.S. labor markets," *American Economic Review*, 103, 2121–2168.

——— (2019): "When work disappears: Manufacturing decline and the falling marriage market value of young men," *American Economic Review: Insights*, 1, 161–78.

Avenevoli, S., J. Swendsen, J.-P. He, M. Burstein, and K. R. Merikangas (2015): "Major depression in the national comorbidity survey–adolescent supplement: Prevalence, correlates, and treatment," *Journal of the American Academy of Child & Adolescent Psychiatry*, 54, 37–44.

Barr, A. and A. Smith (2019): "Fighting Crime in the Cradle: The Effects of Early Childhood Access to Nutritional Assistance," *Working Paper*.

Borusyak, K., X. Jaravel, and J. Spiess (2022): "Revisiting Event Study Designs: Robust and Efficient Estimation," *Working Paper*.

Callaway, B., A. Goodman-Bacon, and P. H. Sant'Anna (2021): "Difference-in-Differences with a Continuous Treatment," *Working Paper*.

Card, D. (1992): "Using regional variation in wages to measure the effects of the federal minimum wage," *Industrial and Labor Relations Review*, 46, 22–37.

Centers for Disease Control and Prevention (2022): "Data and Statistics on Children's Mental Health. Fact Sheet." https://www.cdc.gov/childrensmentalhealth/data.html, Retrieved March 16, 2022.

Chang, F., Y. Jiang, P. Loyalka, J. Chu, Y. Shi, A. Osborn, and S. Rozelle (2019): "Parental migration, educational achievement, and mental health of junior high school students in rural China," *China Economic Review*, 54, 337–349.

Chen, Z.-y., X.-d. Yang, and X.-y. Li (2009): "Psychometric features of CES-D in Chinese adolescents." *Chinese Journal of Clinical Psychology*.

Chinese Academy of Sciences (2020): "Report on National Mental Health Development in China," Chapter 8, Institute of Psychology, Chinese Academy of Sciences.

Costello, E. J., S. Mustillo, A. Erkanli, G. Keeler, and A. Angold (2003): "Prevalence and development of psychiatric disorders in childhood and adolescence," *Archives of General Psychiatry*, 60, 837–844.

Currie, J. (2020): "Child health as human capital," *Health Economics*, 29, 452–463.

Currie, J. and M. Stabile (2006): "Child mental health and human capital accumulation: the case of ADHD," *Journal of Health Economics*, 25, 1094–1118.

——— (2009): "Mental Health in Childhood and Human Capital," in *The Problems of Disadvantaged Youth: An Economic Perspective*, University of Chicago Press.

Currie, J., M. Stabile, P. Manivong, and L. L. Roos (2010): "Child health and young adult outcomes," *Journal of Human Resources*, 45, 517–548.

de Chaisemartin, C. and X. D'Haultfoeuille (2018): "Fuzzy differences-in-differences," *Review of Economic Studies*, 85, 999–1028.

——— (2020): "Two-way fixed effects estimators with heterogeneous treatment effects," *American Economic Review*, 110, 2964–2996.

Dell, M., B. Feigenberg, and K. Teshima (2019): "The violent consequences of trade-induced worker displacement in mexico," *American Economic Review: Insights*, 1, 43–58.

Dix-Carneiro, R. and B. K. Kovak (2019): "Margins of labor market adjustment to trade," *Journal of International Economics*, 117, 125–142.

Dix-Carneiro, R., R. R. Soares, and G. Ulyssea (2018): "Economic shocks and crime: Evidence from the brazilian trade liberalization," *American Economic Journal: Applied Economics*, 10, 158–95.

Erten, B. and P. Keskin (2021): "Trade-offs? The Impact of WTO Accession on Intimate Partner Violence in Cambodia," *forthcoming in Review of Economics and Statistics*.

Erten, B. and J. Leight (2021): "Exporting out of agriculture: The impact of WTO accession on structural transformation in China," *Review of Economics and Statistics*, 103, 364–380.

Feler, L. and M. Z. Senses (2017): "Trade shocks and the provision of local public goods," *American Economic Journal: Economic Policy*, 9, 101–43.

Fletcher, J. M. (2010): "Adolescent depression and educational attainment: results using sibling fixed effects," *Health Economics*, 19, 855–871.

Frisch, R. and F. V. Waugh (1933): "Partial Time Regressions as Compared with Individual Trends," *Econometrica*, 1, 387–401.

Goodman-Bacon, A. (2021): "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*, 225, 254–277.

Greenberger, E., C. Chen, S. R. Tally, and Q. Dong (2000): "Family, peer, and individual correlates of depressive symptomatology among US and Chinese adolescents." *Journal of Consulting and Clinical Psychology*, 68, 209.

Hammen, C., P. A. Brennan, D. Keenan-Miller, and N. R. Herr (2008): "Early onset recurrent subtype of adolescent depression: Clinical and psychosocial correlates," *Journal of Child Psychology and Psychiatry*, 49, 433–440.

Hankin, B. L. (2015): "Depression from childhood through adolescence: Risk mechanisms across multiple systems and levels of analysis," *Current Opinion in Psychology*, 4, 13–20.

Hankin, B. L., L. Y. Abramson, T. E. Moffitt, P. A. Silva, R. McGee, and K. E. Angell (1998): "Development of depression from preadolescence to young adulthood:

emerging gender differences in a 10-year longitudinal study." *Journal of Abnormal Psychology*, 107, 128.

Heckman, J. and R. Pinto (2022): "Econometric Causality," *Forthcoming Annual Reviews of Economics*.

Heckman, J. J. (2006): "Skill formation and the economics of investing in disadvantaged children," *Science*, 312, 1900–1902.

——— (2007): "The economics, technology, and neuroscience of human capability formation," *Proceedings of the National Academy of Sciences*, 104, 13250–13255.

——— (2012): "The developmental origins of health," *Health Economics*, 21, 24.

Heckman, J. J. and R. Pinto (2015): "Causal Analysis after Haavelmo," *Econometric Theory*, 31, 115–151.

Heckman, J. J. and J. Yi (2012): "Human capital, economic growth, and inequality in China," Tech. rep., National Bureau of Economic Research.

Holland, P. W. (1986): "Statistics and Causal Inference," *Journal of the American Statistical Association*, 81, 945–960.

Hu, T.-w., Y. He, M. Zhang, and N. Chen (2007): "Economic costs of depression in China," *Social Psychiatry and Psychiatric Epidemiology*, 42, 110–116.

Jaworska, N. and G. MacQueen (2015): "Adolescence as a unique developmental period," *Journal of Psychiatry & Neuroscience: JPN*, 40, 291.

Keenan-Miller, D., C. L. Hammen, and P. A. Brennan (2007): "Health outcomes related to early adolescent depression," *Journal of Adolescent Health*, 41, 256–262.

Kennard, B. D., G. N. Clarke, V. R. Weersing, J. R. Asarnow, W. Shamseddeen, G. Porta, M. Berk, J. L. Hughes, A. Spirito, G. J. Emslie, et al. (2009): "Effective components of TORDIA cognitive–behavioral therapy for adolescent depression: Preliminary findings." *Journal of Consulting and Clinical Psychology*, 77, 1033.

Kessler, R. C., P. Berglund, O. Demler, R. Jin, K. R. Merikangas, and E. E. Walters (2005): "Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication," *Archives of General Psychiatry*, 62, 593–602.

Khandelwal, A., P. K. Schott, and S.-J. Wei (2013): "Trade liberalization and embedded institutional reform: Evidence from Chinese exporters," *American Economic Review*, 103, 2169–2195.

Khanna, G., E. Murathanoglu, C. Theoharides, and D. Yang (2022): "Abundance from Abroad: Migrant Income and Long-Run Economic Development," *Working Paper*.

Knickmeyer, R. C., S. Gouttard, C. Kang, D. Evans, K. Wilber, J. K. Smith, R. M. Hamer, W. Lin, G. Gerig, and J. H. Gilmore (2008): "A structural MRI study of human brain development from birth to 2 years," *Journal of neuroscience*, 28, 12176–12182.

Knudsen, E. I., J. J. Heckman, J. L. Cameron, and J. P. Shonkoff (2006): "Economic, neurobiological, and behavioral perspectives on building America's future workforce," *Proceedings of the National Academy of Sciences*, 103, 10155–10162.

Kohout, F. J., L. F. Berkman, D. A. Evans, and J. Cornoni-Huntley (1993): "Two shorter forms of the CES-D depression symptoms index," *Journal of Aging and Health*, 5, 179–193.

Lang, M., T. C. McManus, and G. Schaur (2019): "The effects of import competition on health in the local economy," *Health Economics*, 28, 44–56.

Lewandowski, R. E., M. C. Acri, K. E. Hoagwood, M. Olfson, G. Clarke, W. Gardner, S. H. Scholle, S. Byron, K. Kelleher, H. A. Pincus, et al. (2013): "Evidence for the management of adolescent depression," *Pediatrics*, 132, e996–e1009.

Lewinsohn, P. M., P. Rohde, J. R. Seeley, D. N. Klein, and I. H. Gotlib (2003): "Psychosocial functioning of young adults who have experienced and recovered from major depressive disorder during adolescence." *Journal of Abnormal Psychology*, 112, 353.

Li, J.-Y., J. Li, J.-H. Liang, S. Qian, R.-X. Jia, Y.-Q. Wang, and Y. Xu (2019): "Depressive symptoms among children and adolescents in China: A systematic review and meta-analysis," *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*, 25, 7459.

Li, W., Y. Yang, Z.-H. Liu, Y.-J. Zhao, L. Zhang, T. Cheung, C. H. Ng, and Y.-T. Xiang (2021): "The first national action plan on depression in China: Progress and challenges," *The Lancet Regional Health–Western Pacific*, 7.

Lovell, M. (1963): “Seasonal Adjustment of Economic Time Series and Multiple Regression Analysis,” *Journal of the American Statistical Association*, 58, 993–1010.

Malaspina, D., C. Corcoran, K. Kleinhaus, M. Perrin, S. Fennig, D. Nahon, Y. Friedlander, and S. Harlap (2008): “Acute maternal stress in pregnancy and schizophrenia in offspring: a cohort prospective study,” *BMC Psychiatry*, 8, 1–9.

March, J., S. Silva, S. Petrycki, J. Curry, K. Wells, J. Fairbank, B. Burns, M. Domino, S. McNulty, B. Vitiello, et al. (2004): “Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression: Treatment for Adolescents With Depression Study (TADS) randomized controlled trial.” *Jama*, 292, 807–820.

Marmorstein, N. R., W. G. Iacono, and L. Legrand (2014): “Obesity and depression in adolescence and beyond: reciprocal risks,” *International Journal of Obesity*, 38, 906–911.

Mathers, C. D. and D. Loncar (2006): “Projections of global mortality and burden of disease from 2002 to 2030,” *PLoS Medicine*, 3, e442.

McCaig, B. and N. Pavcnik (2018): “Export markets and labor allocation in a low-income country,” *American Economic Review*, 108, 1899–1941.

McManus, T. C. and G. Schaur (2016): “The effects of import competition on worker health,” *Journal of International Economics*, 102, 160–172.

Milkie, M. A., K. M. Nomaguchi, and K. E. Denny (2015): “Does the amount of time mothers spend with children or adolescents matter?” *Journal of Marriage and Family*, 77, 355–372.

National Health Commission of China (2020): “Notice of the general office of the national health commission on the exploration and implementation of special services for the prevention and treatment of depression and Senile dementia (in Chinese),” http://www.nhc.gov.cn/jkj/s7914/202009/a63d8f82eb53451f97217bef0962b98f.shtml, Retrieved March 16, 2022.

National Institute of Mental Health (2022): “Prevalence of Major Depressive Episode Among Adolescents. Fact Sheet.” https://www.nimh.nih.gov/health/statistics/major-depression, Retrieved March 16, 2022.

NUNN, N. (2007): "Relationship-specificity, incomplete contracts, and the pattern of trade," *The Quarterly Journal of Economics*, 122, 569–600.

PAGE, M., J. SCHALLER, AND D. SIMON (2019): "The effects of aggregate and gender-specific labor demand shocks on child health," *Journal of Human Resources*, 54, 37–78.

PERSSON, P. AND M. ROSSIN-SLATER (2018): "Family ruptures, stress, and the mental health of the next generation," *American Economic Review*, 108, 1214–52.

PIERCE, J. AND P. SCHOTT (2016): "The surprisingly swift decline of U.S. manufacturing employment," *American Economic Review*, 106, 1632–1662.

PIERCE, J. R. AND P. K. SCHOTT (2020): "Trade liberalization and mortality: evidence from US counties," *American Economic Review: Insights*, 2, 47–64.

RADLOFF, L. S. (1977): "The CES-D scale: A self-report depression scale for research in the general population," *Applied Psychological Measurement*, 1, 385–401.

RANKIN, S. H., M. E. GALBRAITH, AND S. JOHNSON (1993): "Reliability and validity data for a Chinese translation of the Center for Epidemiological Studies-Depression," *Psychological Reports*, 73, 1291–1298.

RAUCH, J. E. (1999): "Networks versus markets in international trade," *Journal of International Economics*, 48, 7–35.

ROHDE, P., P. M. LEWINSOHN, D. N. KLEIN, J. R. SEELEY, AND J. M. GAU (2013): "Key characteristics of major depressive disorder occurring in childhood, adolescence, emerging adulthood, and adulthood," *Clinical Psychological Science*, 1, 41–53.

RUBIN, D. B. (1978): "Bayesian Inference for Causal Effects: The Role of Randomization," *Annals of Statistics*, 6, 34–58.

RUSHTON, J. L., M. FORCIER, AND R. M. SCHECTMAN (2002): "Epidemiology of depressive symptoms in the National Longitudinal Study of Adolescent Health," *Journal of the American Academy of Child & Adolescent Psychiatry*, 41, 199–205.

RUTTER, M., J. KIM-COHEN, AND B. MAUGHAN (2006): "Continuities and discontinuities in psychopathology between childhood and adult life," *Journal of Child Psychology and Psychiatry*, 47, 276–295.

Steffick, D. E. et al. (2000): "Documentation of affective functioning measures in the Health and Retirement Study," *Ann Arbor, MI: University of Michigan*.

Sun, L. and S. Abraham (2021): "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of Econometrics*, 225, 175–199.

Susser, E., H. W. Hoek, and A. Brown (1998): "Neurodevelopmental disorders after prenatal famine: the story of the Dutch Famine Study," *American Journal of Epidemiology*, 147, 213–216.

Thapar, A., S. Collishaw, D. S. Pine, and A. K. Thapar (2012): "Depression in adolescence," *The Lancet*, 379, 1056–1067.

Tolbert, C. M. and M. Sizer (1996): "US commuting zones and labor market areas: A 1990 update," Tech. rep.

Tong, L., Q. Yan, and I. Kawachi (2019): "The factors associated with being left-behind children in China: Multilevel analysis with nationally representative data," *PloS One*, 14, e0224205.

Wilson, S., B. M. Hicks, K. T. Foster, M. McGue, and W. G. Iacono (2015): "Age of onset and course of major depressive disorder: associations with psychosocial functioning outcomes in adulthood," *Psychological Medicine*, 45, 505–514.

World Health Organization (2017): "Depression and other common mental disorders," Tech. rep., World Health Organization Geneva, Switzerland.

——— (2019): "The WHO special initiative for mental health (2019-2023): universal health coverage for mental health," Tech. rep., World Health Organization.

——— (2021): "WHO menu of cost-effective interventions for mental health," Tech. rep., World Health Organization.

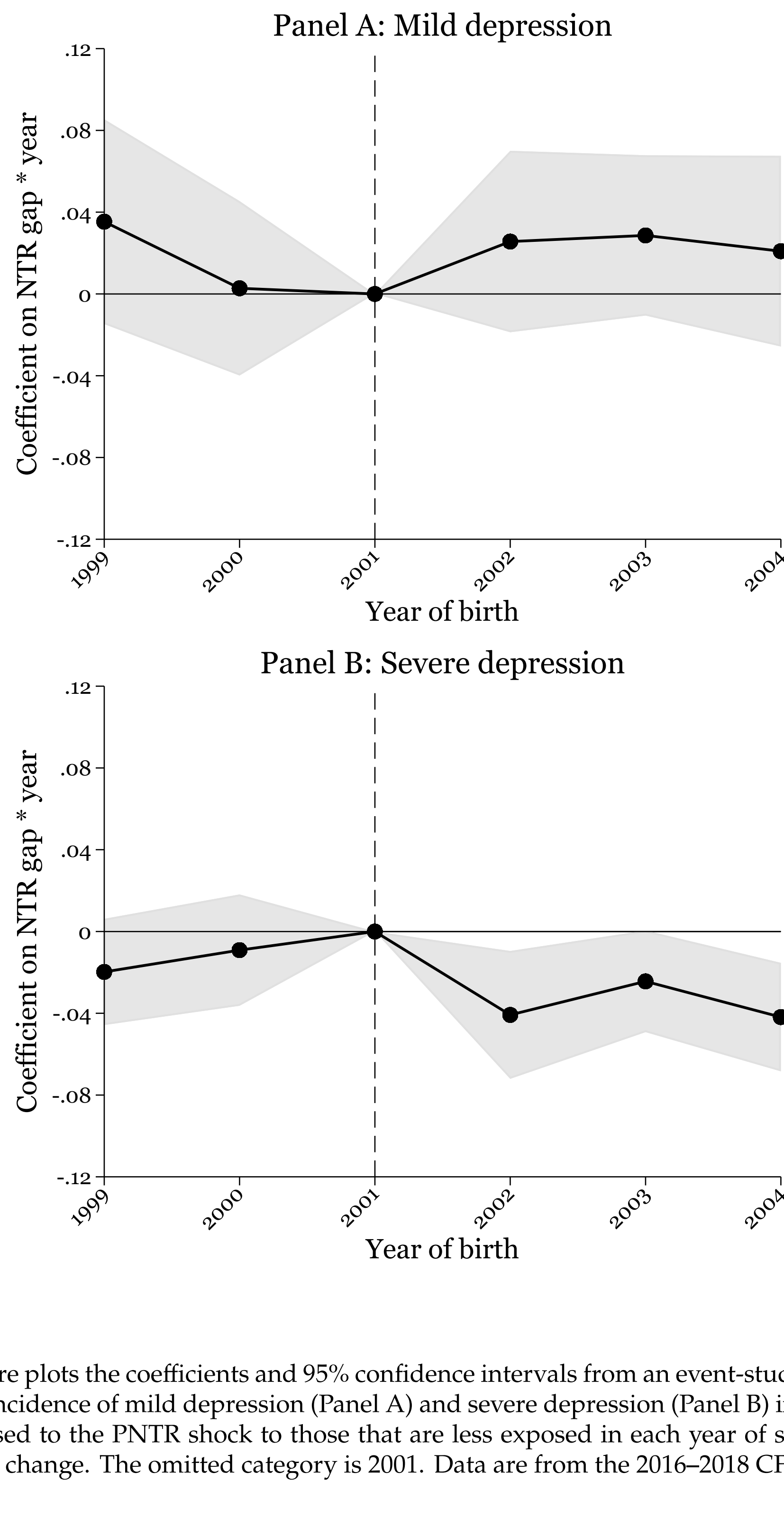
Yitzhaki, S. (2013): *The Gini methodology : a primer on a statistical methodology*, Springer series in statistics, New York: Springer.

Zhang, J. and J. M. Norvilitis (2002): "Measuring Chinese psychological well-being with Western developed instruments," *Journal of Personality Assessment*, 79, 492–511.

Zhou, M., G. Zhang, S. Rozelle, K. Kenny, and H. Xue (2018): "Depressive symptoms of Chinese children: prevalence and correlated factors among subgroups," *International Journal of Environmental Research and Public Health*, 15, 283.

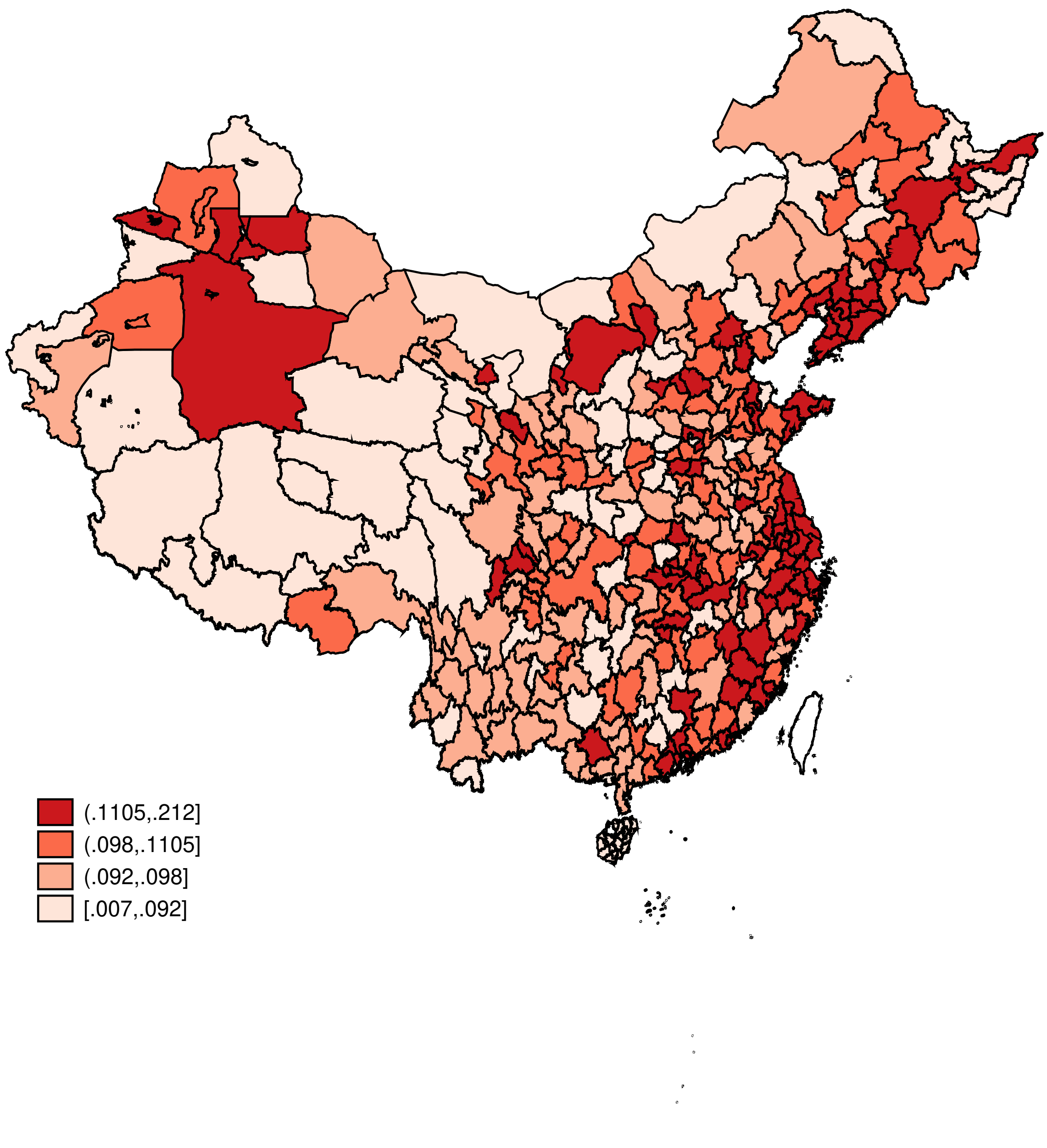
Zisook, S., I. Lesser, J. W. Stewart, S. R. Wisniewski, G. Balasubramani, M. Fava, W. S. Gilmer, T. R. Dresselhaus, M. E. Thase, A. A. Nierenberg, et al. (2007): "Effect of age at onset on the course of major depressive disorder," *American Journal of Psychiatry*, 164, 1539–1546.

Figure 1: Event Study: Exposure to PNTR and Severe Depression Incidence



Note: This figure plots the coefficients and 95% confidence intervals from an event-study regression that compares the incidence of mild depression (Panel A) and severe depression (Panel B) in prefectures that are more exposed to the PNTR shock to those that are less exposed in each year of survey before and after the policy change. The omitted category is 2001. Data are from the 2016–2018 CFPS.

Figure 2: Prefecture-Level Exposure to PNTR



Note: This figure plots the prefecture-level exposure to PNTR, computed as the employment-share weighted-average NTR gap across all of the Chinese three-digit industries in 1999. Employment data are from the 1990 population census. Data on industry-level NTR gap are from Pierce and Schott (2016).

Table 1A: Summary Statistics

	Obs (1)	Mean (2)	S.D. (3)	Min (4)	Max (5)
Panel A: CFPS 2016-2018: Cohorts born between 1999 and 2004					
Health and cognition outcomes					
Mild depression	3,549	0.156	0.363	0.000	1.000
Severe depression	3,549	0.059	0.237	0.000	1.000
Mental health index	3,549	0.000	1.000	-6.642	1.300
Physical health index	3,547	0.000	1.000	-4.981	0.433
Cognitive function index	2,984	0.000	1.000	-7.385	5.042
Demographic characteristics					
Age	3,549	15.413	1.999	12.000	19.000
Male	3,549	0.526	0.499	0.000	1.000
Father completed middle school	3,549	0.516	0.500	0.000	1.000
Mother completed middle school	3,549	0.394	0.489	0.000	1.000
Panel B: CHNS 2000-2015: Cohorts born between 1999 and 2004					
Pregnancy History File: early life investment					
Months of breast-feeding	409	10.460	5.572	0.000	36.000
Number of vaccinations	543	1.967	2.374	0.000	9.000
BCG vaccination	543	0.201	0.401	0.000	1.000
Hepatitis B vaccination	543	0.269	0.444	0.000	1.000
DPT vaccination	543	0.271	0.445	0.000	1.000
Encephalities B vaccination	543	0.234	0.424	0.000	1.000
Measles vaccination	543	0.234	0.424	0.000	1.000
Polio vaccination	543	0.346	0.476	0.000	1.000
Child sample: nutrition intake in past 3 days (children ages 0-12)					
Calories	1,650	1310.917	534.650	238.167	4646.358
Protein	1,650	43.211	20.173	5.762	174.500
Carbohydrate	1,650	178.554	76.154	32.331	598.076
Fat	1,650	47.003	30.122	1.482	422.754
Child sample: child development (children ages 0-12)					
Height (cm)	1,876	120.553	24.662	50.000	172.800
Weight (kg)	1,928	25.7944	13.019	3.000	110.000
Childcare by people outside the household (children ages 0-6)					
Hours per day in a typical day	676	3.120	4.770	0.000	24.000
Days per week in a typical week	795	1.774	2.532	0.000	7.000
Demographic characteristics (children ages 0-12)					
Age	1,650	6.600	2.884	0.000	12.000
Male	1,650	0.543	0.498	0.000	1.000
Father completed middle school	1,650	0.759	0.428	0.000	1.000
Mother completed middle school	1,650	0.675	0.469	0.000	1.000

Notes: Panels A and B present the summary statistics for child development variables and demographic characteristics from the 2016 and 2018 CFPS sample and from the 2000, 2004, 2006, 2009, 2011 and 2015 CHNS sample, respectively. All variables are summarized at the child level.

Table 1B: Summary Statistics

	Obs (1)	Mean (2)	S.D. (3)	Min (4)	Max (5)
Panel C: CHNS 1991-2015					
Pregnant women during the sample period of survey					
Prenatal visits	1,958	5.001	4.976	0.000	78.000
Panel D: City Statistical Yearbook 1995-2015					
Labor income (in log)	5,628	13.039	1.238	4.736	18.222
Labor income per worker (in log)	5,628	8.419	2.555	-0.085	14.233
Panel E: Census 1990-2015					
Female sample: share of margins of labor market adjustment to females aged 20-55					
Total employment	1,655	0.766	0.127	0.251	0.990
Employment in agriculture	1,655	0.461	0.236	0.000	0.981
Employment in manufacturing	1,655	0.100	0.099	0.000	0.713
Employment in service	1,655	0.204	0.104	0.004	0.778
Unemployed	1,655	0.035	0.029	0.000	0.222
Not in the labor force	1,655	0.200	0.110	0.010	0.746
Male sample: share of margins of labor market adjustment to males aged 22-60					
Total employment	1,655	0.909	0.055	0.603	0.990
Employment in agriculture	1,655	0.485	0.217	0.003	0.961
Employment in manufacturing	1,655	0.120	0.094	0.000	0.695
Employment in service	1,655	0.304	0.130	0.024	0.898
Unemployed	1,655	0.027	0.024	0.000	0.205
Not in the labor force	1,655	0.065	0.036	0.008	0.306

Notes: Panel C presents the summary statistics of the number of prenatal visits from the 1991, 1993, 1997, 2000, 2004, 2006, 2009, 2011 and 2015 CHNS pregnant women sample. Panel D presents the summary statistics of labor income and labor income per worker (in log) from the 1995-2015 City Statistical Yearbook. Panel E presents the summary statistics of employment variables separately for female working age population (ages 20 to 55 years) and male working age population (ages 22 to 60 years) from the 1990, 2000, 2005, 2010, and 2015 population census sample. All variables are summarized at the prefecture level.

Table 2: Impact of PNTR on Adolescent Mental Health Outcomes

	(1)	(2)	(3)
Panel A: Mild depression			
Post × NTR gap	0.012	0.009	0.011
	(0.013)	(0.012)	(0.012)
Observations	3549	3549	3549
Outcome mean	0.16	0.16	0.16
Panel B: Severe depression			
Post × NTR gap	-0.026***	-0.026***	-0.025***
	(0.008)	(0.009)	(0.009)
Observations	3549	3549	3549
Outcome mean	0.06	0.06	0.06
Birth prefecture fixed effects	Yes	Yes	Yes
Birth year fixed effects	Yes	Yes	Yes
Survey year fixed effects	Yes	Yes	Yes
Birth year fixed effects			
× Other trade policies	Yes	Yes	Yes
× Initial prefecture characteristics		Yes	Yes
Individual characteristics			Yes

Notes: Data are from the 2016–2018 CFPS. This table reports results of the difference-in-difference regressions of adolescent mental health outcomes on interaction of the prefecture-level NTR gap and a post-PNTR indicator. Regressions in column 1 control for birth year fixed effects, prefecture of birth fixed effects, survey year fixed effects, and the birth year fixed effects interacted with other trade policies including China's output tariffs, input tariffs, the NTR rates, export licensing, MFA quotas, and contract intensity. Regressions in column 2 further control for the birth year fixed effects interacted with initial prefecture characteristics including GDP per capita, share of employment in manufacturing, and distance to the nearest port. Regressions in column 3 further control for individual characteristics including adolescent age fixed effects, adolescent gender, and indicator variables for whether the mother and father completed middle school. Standard errors are clustered at the prefecture of birth level. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

Table 3: Impact of PNTR on Labor Income

	Log labor income		Log labor income per worker	
	(1)	(2)	(3)	(4)
Post × NTR gap	0.056***	0.034**	0.101***	0.065***
	(0.015)	(0.015)	(0.016)	(0.013)
Observations	5628	5628	5628	5628
Outcome mean	13.04	13.04	8.42	8.42
Prefecture fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Year fixed effects				
× Other trade policies	Yes	Yes	Yes	Yes
× Initial prefecture characteristics		Yes		Yes

Notes: Data are from the 1995–2015 China City Statistical Yearbooks. This table reports results of the difference-in-difference regressions of labor income and labor income per worker (in log) on interaction of the prefecture-level NTR gap and a post-PNTR indicator. Regressions in columns 1 and 3 control for prefecture fixed effects, year fixed effects, and the year fixed effects interacted with other trade policies including China's output tariffs, input tariffs, the NTR rates, export licensing, MFA quotas, and contract intensity. Regressions in columns 2 and 4 further control for the year fixed effects interacted with initial prefecture characteristics including GDP per capita, share of employment in manufacturing, and distance to the nearest port. Standard errors are clustered at the prefecture level. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

Table 4: Impact of PNTR on Early Life Investments

	Prenatal visits (1)	Months of breastfeeding (2)	Number of vaccinations (3)	BCG vaccination (4)	Hepatitis B vaccination
Post × NTR gap	1.112**	1.951**	0.628***	0.113***	0.101**
	(0.509)	(0.764)	(0.193)	(0.041)	(0.043)
Observations	1958	409	543	543	543
Outcome mean	5.01	10.46	1.97	0.20	0.27
Prefecture fixed effects	Yes	Yes	Yes	Yes	Yes
Birth year fixed effects		Yes	Yes	Yes	Yes
Survey year fixed effects	Yes	Yes	Yes	Yes	Yes
Birth year fixed effects					
× Other trade policies	Yes	Yes	Yes	Yes	Yes
× Initial prefecture characteristics	Yes	Yes	Yes	Yes	Yes
Individual characteristics	Yes	Yes	Yes	Yes	Yes
	DPT vaccination (6)	Encephalitis B vaccination (7)	Measles vaccination (8)	Polio vaccination (9)	
Post × NTR gap	0.197***	0.073**	0.101***	0.056	
	(0.029)	(0.034)	(0.034)	(0.068)	
Observations	543	543	543	543	
Outcome mean	0.27	0.23	0.23	0.35	
Prefecture fixed effects	Yes	Yes	Yes	Yes	
Birth year fixed effects	Yes	Yes	Yes	Yes	
Survey year fixed effects	Yes	Yes	Yes	Yes	
Birth year fixed effects					
× Other trade policies	Yes	Yes	Yes	Yes	
× Initial prefecture characteristics	Yes	Yes	Yes	Yes	
Individual characteristics	Yes	Yes	Yes	Yes	

Notes: Prenatal visits information comes from data of pregnant women during the sample period of 1991-2015 CHNS. Column (1) regression controls for prefecture fixed effects, survey year fixed effects, survey year fixed effects interacted with initial prefecture characteristics, survey year fixed effects interacted with other trade policies, and the individual characteristics including the mother's age fixed effects and an indicator variable for whether the mother completed middle school. The information on breastfeeding duration and vaccinations reported in columns (2)–(9) comes from data on children from the 2000-2015 CHNS Pregnancy History File. Regressions control for the child's birth year fixed effects, prefecture fixed effects, survey year fixed effects, birth year fixed effects interacted with initial prefecture characteristics including GDP per capita, share of employment in manufacturing, and distance to the nearest port, birth year fixed effects interacted with other trade policies including China's output tariffs, input tariffs, the NTR rates, export licensing, MFA quotas, and contract intensity, and individual characteristics including the child's age fixed effects, gender, and indicator variables for whether the mother and father completed middle school. Standard errors are clustered at the prefecture level. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

Table 5: Impact of PNTR on Nutrition Intake and Child Development Outcomes

	(1)	(2)	(3)
Panel A: Total Calories			
Post × NTR gap	75.603***	55.073**	59.452***
	(19.535)	(21.986)	(20.732)
Observations	1650	1650	1650
Outcome mean	1310.92	1310.92	1310.92
Panel B: Protein			
Post × NTR gap	3.504***	2.923***	2.664***
	(0.624)	(0.679)	(0.723)
Observations	1650	1650	1650
Outcome mean	43.21	43.21	43.21
Panel C: Carbohydrate			
Post × NTR gap	8.799***	6.103*	6.203*
	(2.762)	(3.193)	(3.117)
Observations	1650	1650	1650
Outcome mean	178.55	178.55	178.55
Panel D: Fat			
Post × NTR gap	2.965***	2.146	2.705*
	(1.061)	(1.574)	(1.519)
Observations	1650	1650	1650
Outcome mean	47.00	47.00	47.00
Panel E: Height			
Post × NTR gap	0.645	0.833*	0.933**
	(0.389)	(0.487)	(0.461)
Observations	1876	1876	1876
Outcome mean	120.55	120.55	120.55
Panel F: Weight			
Post × NTR gap	1.059**	1.999***	1.318**
	(0.435)	(0.574)	(0.501)
Observations	1928	1928	1928
Outcome mean	25.79	25.79	25.79
Prefecture fixed effects	Yes	Yes	Yes
Birth year fixed effects	Yes	Yes	Yes
Survey year fixed effects	Yes	Yes	Yes
Birth year fixed effects			
× Other trade policies	Yes	Yes	Yes
× Initial prefecture characteristics		Yes	Yes
Individual characteristics			Yes

Notes: Data are from the 2000–2015 CHNS. This table reports results of the difference-in-difference regressions of nutrition intake on interaction of the prefecture-level NTR gap and a post-PNTR indicator. Regression in column 1 controls for birth year fixed effects, prefecture fixed effects, survey year fixed effects, and the birth year fixed effects interacted with other trade policies including China's output tariffs, input tariffs, the NTR rates, export licensing, MFA quotas, and contract intensity. Regression in column 2 further controls for the birth year fixed effects interacted with initial prefecture characteristics including GDP per capita, share of employment in manufacturing, and distance to the nearest port. Regression in column 3 further controls for individual characteristics including child age fixed effects, child gender, and indicator variables for whether the mother and father completed middle school. Standard errors are clustered at the prefecture of birth level. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

Table 6: Impact of PNTR on Female and Male Employment

	All sectors (1)	Agri (2)	Manu (3)	Service (4)	Unemployed (5)	NILF (6)
Panel A: Female						
Post × NTR gap	-0.010***	-0.050***	0.018***	0.021***	-0.000	0.011***
	(0.004)	(0.006)	(0.003)	(0.003)	(0.001)	(0.004)
Observations	1655	1655	1655	1655	1655	1655
Outcome mean	0.77	0.46	0.10	0.20	0.03	0.20
Panel B: Male						
Post × NTR gap	0.004**	-0.043***	0.017***	0.030***	-0.000	-0.004***
	(0.002)	(0.005)	(0.004)	(0.004)	(0.001)	(0.001)
Observations	1655	1655	1655	1655	1655	1655
Outcome mean	0.91	0.48	0.12	0.30	0.03	0.06
Prefecture fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects						
× Other trade policies	Yes	Yes	Yes	Yes	Yes	Yes
× Initial prefecture characteristics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Data are from the 1990, 2000, 2005, 2010, and 2015 population census in China. This table reports results of the difference-in-difference regressions of margins of local labor market adjustment on interaction of the prefecture-level NTR gap and a post-PNTR indicator. All regressions control for prefecture fixed effects, year fixed effects, the year fixed effects interacted with other trade policies including China's output tariffs, input tariffs, the NTR rates, export licensing, MFA quotas, and contract intensity, and the survey year fixed effects interacted with initial prefecture characteristics including GDP per capita, share of employment in manufacturing, and distance to the nearest port. Regressions in Panel A consider marriageable and working-age women (ages 20-55). Regressions in Panel B consider marriageable and working-age men (ages 22-60). Regressions are weighted by 1990 prefecture population. Standard errors are clustered at the prefecture level. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

Table 7: Impact of PNTR on Childcare Provision

	(1)	(2)	(3)
Panel A: Hours per day cared by people outside the household in a typical day			
Post × NTR gap	-0.870***	-1.278***	-1.332***
	(0.305)	(0.359)	(0.357)
Observations	676	676	676
Outcome mean	3.12	3.12	3.12
Panel B: Days per week cared by people outside the household in a typical week			
Post × NTR gap	-0.758***	-0.963***	-1.014***
	(0.258)	(0.257)	(0.232)
Observations	795	795	795
Outcome mean	1.77	1.77	1.77
Prefecture fixed effects	Yes	Yes	Yes
Birth year fixed effects	Yes	Yes	Yes
Survey year fixed effects	Yes	Yes	Yes
Birth year fixed effects			
× Other trade policies	Yes	Yes	Yes
× Initial prefecture characteristics		Yes	Yes
Individual characteristics			Yes

Notes: Data are from the 2000–2015 CHNS. This table reports results of the difference-in-difference regressions of childcare provision on interaction of the prefecture-level NTR gap and a post-PNTR indicator. Regressions in column 1 control for birth year fixed effects, prefecture fixed effects, survey year fixed effects, and the birth year fixed effects interacted with other trade policies including China's output tariffs, input tariffs, the NTR rates, export licensing, MFA quotas, and contract intensity. Regressions in column 2 further control for the birth year fixed effects interacted with initial prefecture characteristics including GDP per capita, share of employment in manufacturing, and distance to the nearest port. Regressions in column 3 further control for individual characteristics including child age fixed effects, child gender, and indicator variables for whether the mother and father completed middle school. Standard errors are clustered at the prefecture of birth level. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

Table 8: Impact of PNTR on Parenting Styles

	(1)	(2)	(3)
Panel A: Parenting style- Optimal			
Post × NTR gap	0.040***	0.041***	0.044***
	(0.014)	(0.015)	(0.015)
Observations	3149	3149	3149
Outcome mean	0.16	0.16	0.16
Panel B: Parenting style- Affectionate			
Post × NTR gap	-0.014	-0.028	-0.030
	(0.025)	(0.020)	(0.021)
Observations	3149	3149	3149
Outcome mean	0.34	0.34	0.34
Panel C: Parenting style- Affectionless			
Post × NTR gap	0.013	0.010	0.010
	(0.016)	(0.017)	(0.016)
Observations	3149	3149	3149
Outcome mean	0.17	0.17	0.17
Panel D: Parenting style- Neglectful			
Post × NTR gap	-0.039*	-0.024	-0.024
	(0.021)	(0.018)	(0.019)
Observations	3149	3149	3149
Outcome mean	0.32	0.32	0.32
Birth prefecture fixed effects	Yes	Yes	Yes
Birth year fixed effects	Yes	Yes	Yes
Survey year fixed effects	Yes	Yes	Yes
Birth year fixed effects			
× Other trade policies	Yes	Yes	Yes
× Initial prefecture characteristics		Yes	Yes
Individual characteristics			Yes

Notes: Data are from the 2016–2018 CFPS. This table reports results of the difference-in-difference regressions of adolescent mental health outcomes on interaction of the prefecture-level NTR gap and a post-PNTR indicator. Regressions in column 1 control for birth year fixed effects, prefecture of birth fixed effects, survey year fixed effects, and the birth year fixed effects interacted with other trade policies including China's output tariffs, input tariffs, the NTR rates, export licensing, MFA quotas, and contract intensity. Regressions in column 2 further control for the birth year fixed effects interacted with initial prefecture characteristics including GDP per capita, share of employment in manufacturing, and distance to the nearest port. Regressions in column 3 further control for individual characteristics including adolescent age fixed effects, adolescent gender, and indicator variables for whether the mother and father completed middle school. Standard errors are clustered at the prefecture of birth level. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

FOR ONLINE PUBLICATION

Appendix A Mathematical Proofs

A.1 Proof of Proposition P.1

The proposition requires multiple applications of the formula for the integration by parts. Namely, let $h(x):\mathbb{R}\to\mathbb{R}$ be an integrable function and $g(x):\mathbb{R}\to\mathbb{R}$ be an a differentiable function, then the following equation holds:

$$\int_a^b \frac{\partial g(x)}{\partial x} h(x)dx = \Big[g(x)h(x)\Big]_a^b - \int_a^b g(x)\frac{\partial h(x)}{\partial x}dx \tag{44}$$

Note that the moderator M is a continuous random variable in $[\underline{m},\overline{m}]$, and $ATE_t(m|m)$ is differentiable. Thereby $ATE_t(m|m) = E(Y_t(1,m) - Y_t(0,m)|M = m)$ is continuous in $[\underline{m},\overline{m}]$. Moreover, we have that

$$ATE_t = \int_{\underline{m}}^{\overline{m}} ATE_t(m|m) f_M(m)dm,$$

where $f_M(m) > 0$ is the probability density of M and $F_M(m) = \int_{\underline{m}}^{m} f_M(m)dm = P(M \leq m)$ denotes the cumulative probability function of M such that $F_M(\underline{m}) = 0$ and $F_M(\overline{m}) = 1$.

Let m^* be any value in $[\underline{m},\overline{m}]$. We first apply (44) to the integral $\int_{m^*}^{\overline{m}} MATE_t(m)\,(1 - F_M(m))\,dm$:

$$\begin{aligned}
&\int_{m^*}^{\overline{m}} MATE_t(m)\,(1 - F_M(m))\,dm \\
&= \int_{m^*}^{\overline{m}} \frac{\partial ATE_t(m|m)}{\partial m}\,(1 - F_M(m))\,dm \\
&= \Big[ATE_t(m|m)\,(1 - F_M(m))\Big]_{m^*}^{\overline{m}} + \int_{m^*}^{\overline{m}} ATE_t(m|m) f_M(m)dm \\
&= \Big(ATE_t(\overline{m}|\overline{m})\,(1 - F_M(\overline{m}))\Big) - \Big(ATE_t(m^*|m^*)\,(1 - F_M(m^*))\Big) + \int_{m^*}^{\overline{m}} ATE_t(m|m) f_M(m)dm, \\
&= -\Big(ATE_t(m^*|m^*)\cdot(1 - F_M(m^*))\Big) + \int_{m^*}^{\overline{m}} ATE_t(m|m) f_M(m)dm, \\
\therefore &\int_{m^*}^{\overline{m}} ATE_t(m|m) f_M(m)dm = \int_{m^*}^{\overline{m}} MATE_t(m)\,(1 - F_M(m))\,dm + ATE_t(m^*|m^*)\cdot(1 - F_M(m^*))\,,
\end{aligned} \tag{45}$$

where the first equality comes from the definition of $MATE_t(m)$. The second equality applies the integration by parts. The fourth equality is due to the fact that $F_M(\overline{m}) = 1$. The last equality simply rearranges the terms.

Next, we apply (44) to the integral $\int_{\underline{m}}^{m^*} MATE_t(m)\big(-F_M(m)\big)dm$:

$$\begin{aligned}
&\int_{\underline{m}}^{m^*} MATE_t(m)\big(-F_M(m)\big)dm \\
&\quad= \int_{\underline{m}}^{m^*} \frac{\partial ATE_t(m|m)}{\partial m}\big(-F_M(m)\big)dm \\
&\quad= \Big[ATE_t(m|m)\,(-F_M(m))\Big]_{\underline{m}}^{m^*} + \int_{\underline{m}}^{m^*} ATE_t(m|m)f_M(m)dm \\
&\quad= \Big(ATE_t(m^*|m^*)\,(-F_M(m^*))\Big) - \Big(ATE_t(\underline{m}|\underline{m})\,(-F_M(\underline{m}))\Big) + \int_{\underline{m}}^{m^*} ATE_t(m|m)f_M(m)dm, \\
&\quad= \Big(ATE_t \cdot F_M(m^*)\Big) + \int_{\underline{m}}^{m^*} ATE_t(m|m)f_M(m)dm, \\
&\therefore \int_{\underline{m}}^{m^*} ATE_t(m|m)f_M(m)dm = \int_{\underline{m}}^{m^*} MATE_t(m)\big(-F_M(m)\big)dm - ATE_t(m^*|m^*)\cdot F_M(m^*),
\end{aligned} \tag{46}$$

where the first equality comes from the definition of $MATE_t(m)$. The second equality applies the integration by parts. The fourth equality is due to the fact that $F_M(\overline{m}) = 1$. The last equality simply rearranges the terms.

The final expression is obtained by summing equations (45) and (46). The sum of the left-hand side of these two equations give us the average treatment effect:

$$\int_{\underline{m}}^{m^*} ATE_t(m|m)f_M(m)dm + \int_{m^*}^{\overline{m}} ATE_t(m|m)f_M(m)dm = \int_{\underline{m}}^{\overline{m}} ATE_t(m|m)f_M(m)dm = ATE_t. \tag{47}$$

The sum of the right-hand side of equations (45) and (46) give us the following expression:

$$\begin{aligned}
&\left(\int_{\underline{m}}^{m^*} MATE_t(m)\big(-F_M(m)\big)dm - ATE_t(m^*|m^*)\cdot F_M(m^*)\right) \\
&+\left(\int_{m^*}^{\overline{m}} MATE_t(m)\,(1-F_M(m))\,dm + ATE_t(m^*|m^*)\cdot(1-F_M(m^*))\right) \qquad (48)\\
&= \int_{\underline{m}}^{\overline{m}} MATE_t(m)\cdot\Big(\mathbf{1}[m \geq m^*](1-F_M(m)) - \mathbf{1}[m \leq m^*]\cdot F_M(m)\Big)dm + ATE_t(m^*|m^*). \qquad (49)
\end{aligned}$$

We can equate the left-hand in (47) with the right-hand side in (49) to obtain the desired expression:

$$ATE_t = \int_{\underline{m}}^{\overline{m}} MATE_t(m)\cdot\Big(\mathbf{1}[m \geq m^*](1-F_M(m)) - \mathbf{1}[m \leq m^*]\cdot F_M(m)\Big)dm + ATE_t(m^*|m^*).$$

Equation (11) of the proposition arises because, according to Assumption **A.2**, $m_0 \in \mathcal{M}_0 \Rightarrow Y_{it}(1, m_0) = Y_{it}(0, m_0)$ for all units $i \in \mathcal{I}$. Thus $E(Y_t(1) = Y_t(0)|M = m_0) = E_{\mathcal{I}}(Y_{it}(1, m_0) = Y_{it}(0, m_0)|M = m_0) = 0$. Thus $ATE_t(m_0|m_0) = E()0$ for all $m_0 \in \mathcal{M}_0$. Equation (12) of the proposition arises because, if $\underline{m} \in \mathcal{M}_0$, then $ATE_t(\underline{m}|\underline{m}) = 0$ and

$\mathbf{1}[m > m_0] = 1$ for all $m \in (\underline{m}, \overline{m}]$.

The equations also hold if we condition on $D = 1$. In this case, we need to replace ATE_t, $MATE_t$, $ATE_t(m^*|m^*)$, $F_M(m)$ by ATT_t, $MATT_t$, $ATT_t(m^*|m^*)$, $F_{M|D=1}(m)$ respectively. In particular, note that if $m_0 \in \mathcal{M}_0 \Rightarrow Y_{it}(1, m_0) = Y_{it}(0, m_0)$ for all units $i \in \mathcal{I}$, then $ATT_t = E(Y_t(1) = Y_t(0)|M = m_0, D = 1) = 0$.

A.2 Proof of Theorem T.1

The identification of $ATT_t(m|m)$ in (13) is obtained by the following equations:

$$
\begin{aligned}
ATT_t(m|m) &= E[Y_t(1,m) - Y_t(0,m)|D = 1, M = m], \\
&= E[Y_t(1,m) - Y_{t-1}(0,m)|D = 1, M = m] - E[Y_t(0,m) - Y_{t-1}(0,m)|D = 1, M = m], \\
&= E[Y_t(1,m) - Y_{t-1}(0,m)|D = 1, M = m] - E[Y_t(0,m) - Y_{t-1}(0,m)|D = 0, M = m], \\
&= E[\Delta Y_t|D = 1, M = m] - E[\Delta Y_t|D = 0, M = m],
\end{aligned}
$$

where the first equality is due to the definition of $ATT_t(m|m)$ in (5). The second equality adds and subtracts $E[Y_{t-1}(0,m)|D = 1, M = m]$. The third equality invokes the Conditional Parallel Trends Assumption **A.4**. The last equality is due to **A.1**. Namely, the expected value of $Y_t(1,m)$ and $Y_{t-1}(0,m)$ are observed when conditioning on $(D = 1, M = m)$, and the expected value of $Y_t(0,m)$ and $Y_{t-1}(0,m)$ are observed when conditioning on $D = 0, M = m$.

The identification equation for ATT_t in (14) stems from the following equations:

$$
\begin{aligned}
ATT_t &= E[Y_t(1) - Y_t(0)|D = 1], \\
&= \int_m E[Y_t(1) - Y_t(0)|D = 1, M = m] dF_{M|D=1}(m), \\
&= \int_m E[Y_t(1,m) - Y_t(0,m)|D = 1, M = m] dF_{M|D=1}(m), \\
&= \int_m E[Y_t(1,m) - Y_t(0,m)|D = 1, M = m] \frac{P(D = 1|M)}{P(D = 1)} dF_M(m),
\end{aligned}
$$

where the second equation is due to the law of iterated expectations. The third equation is due to **A.1** and the fourth equation is due to the Bayes theorem.

A.3 Proof of Proposition P.3

The proposition describe a non-identification result. It is useful to rewrite the conditional average treatment effect $ATE_t(m|m)$ in terms of the conditional treatment on the treated

$ATT_t(m|m)$:

$$
\begin{aligned}
ATE_t(m|m) &= E(Y_t(1,m) - Y_t(0,m)|M = m) && (50)\\
&= E(Y_t(1,m) - Y_t(0,m)|M = m, D = 1)P(D = 1|M = m)\\
&\quad + E(Y_t(1,m) - Y_t(0,m)|M = m, D = 0)P(D = 0|M = m) && (51)\\
&= ATT_t(m|m)P(D = 1|M = m)\\
&\quad + E(Y_t(1,m) - Y_t(0,m)|M = m, D = 0)P(D = 0|M = m). && (52)
\end{aligned}
$$

The Conditional Parallel Trends Assumption **A.4** enable us to identify $ATT_t(m|m)$, but not $E(Y_t(1,m) - Y_t(0,m)|M = m, D = 0)$, which is the causal effect the treatment for the control group. Note that the control group never experience the treatment itself. Its identification requires an assumption that enable us to use the treatment group to evaluate the causal effect for the control group. The same rationale can be applied to show that the average treatment effect ATE_t is not identified either.

A.4 Proof of Theorem T.2

The identification of $ATE_t(m|m)$ in (16) is obtained by the following equations:

$$
\begin{aligned}
ATE_t(m|m) &= E[Y_t(1,m) - Y_t(0,m)|M = m],\\
&= E[Y_t(1,m) - Y_{t-1}(0,m)|M = m] - E[Y_t(0,m) - Y_{t-1}(0,m)|M = m],\\
&= E[Y_t(1,m) - Y_{t-1}(0,m)|D = 1, M = m] - E[Y_t(0,m) - Y_{t-1}(0,m)|D = 0, M = m],\\
&= E[\Delta Y_t|D = 1, M = m] - E[\Delta Y_t|D = 0, M = m],
\end{aligned}
$$

where the first equality is due to the definition of $ATE_t(m|m)$ in (5). The second equality adds and subtracts $E[Y_{t-1}(0,m)|M = m]$. The third equality invokes the Strong Parallel Trends **A.5** and the Full Support **A.3** . The last equality is due to **A.1**. Namely, the expected value of $Y_t(1,m)$ and $Y_{t-1}(0,m)$ are observed when conditioning on $(D = 1, M = m)$, and the expected value of $Y_t(0,m)$ and $Y_{t-1}(0,m)$ are observed when conditioning on $D = 0, M = m$.

The identification equation for ATE_t in (17) stems from the following equations:

$$
\begin{aligned}
ATE_t &= E[Y_t(1) - Y_t(0)],\\
&= \int_m E[Y_t(1) - Y_t(0)|M = m]dF_M(m),\\
&= \int_m E[Y_t(1,m) - Y_t(0,m)|D = 1, M = m]dF_M(m),
\end{aligned}
$$

where the second equation is due to the law of iterated expectations and the third equation is due to **A.1**.

A.5 Proof of Proposition P.2

The proposition so proved by the following equations:

$$
\begin{aligned}
& ATT_t(m|m) - ATT_t(m'|m') && (53)\\
&= E(Y_t(1,m) - Y_t(0,m)|D=1, M=m) - E(Y_t(1,m') - Y_t(0,m')|D=1, M=m') && (54)\\
&= \underbrace{E(Y_t(1,m) - Y_t(0,m)|D=1, M=m) - E(Y_t(1,m') - Y_t(0,m')|D=1, M=m)}_{ATT_t(m|m) - ATT_t(m'|m)} && (55)\\
&+ \underbrace{E(Y_t(1,m') - Y_t(0,m')|D=1, M=m) - E(Y_t(1,m') - Y_t(0,m')|D=1, M=m')}_{ATT_t(m'|m) - ATT_t(m'|m')}. && (56)
\end{aligned}
$$

A.6 Proof of Theorem T.3

This proof adopts the notation of the main paper: (1) $\overline{M}_d = E(M|D = d); d =\in \{1, 0\}$ denotes the expected value of the moderator condition on the treatment group; (2) $\overline{\Delta Y}_d = E(Y_t - Y_{t-1}|D = d); d =\in \{1, 0\}$ denotes the expected value of the outcome time difference condition on the treatment groups; (3) $P_d = P(D = d); d \in \{0, 1\}$ denotes the probability of each treatment group; and (4) $\overline{\Delta Y} = E(\Delta Y) = \overline{\Delta Y}_1 P_1 + \overline{\Delta Y}_0 P_0$.

The expected value of the OLS estimator of the parameter β_{DiD} in (18) evaluates the following ratio:

$$\beta_{DiD} + \frac{\text{Cov}(\Delta Y_t, D \cdot M)}{\text{Var}(D \cdot M)} \tag{57}$$

We can express the numerator of (57) as:

$$
\begin{aligned}
\text{Cov}(\Delta Y_t, D \cdot M) &= E((\Delta Y_t - \overline{\Delta Y}) \cdot M|D = 1)P_1 && (58)\\
&= E(\Delta Y_t \cdot M|D = 1)P_1 - \overline{\Delta Y} \cdot \overline{M}_1 P_1 && (59)\\
&= E(\Delta Y_t \cdot M|D = 1)P_1 - (\overline{\Delta Y}_1 P_1 + \overline{\Delta Y}_0 P_0)\overline{M}_1 P_1 && (60)\\
&= (E(\Delta Y_t \cdot M|D = 1) - \overline{\Delta Y}_1 P_1 \overline{M}_1 + \overline{\Delta Y}_0 P_0 \overline{M}_1)P_1 && (61)\\
&= (E(\Delta Y_t \cdot M|D = 1) - \overline{\Delta Y}_1 \overline{M}_1 + \overline{\Delta Y}_1 \overline{M}_1(1 - P_1) - \overline{\Delta Y}_0 P_0 \overline{M}_1)P_1 && (62)\\
&= (\text{Cov}(\Delta Y_t, M|D = 1) + \overline{\Delta Y}_1 \overline{M}_1 P_0 - \overline{\Delta Y}_0 P_0 \overline{M}_1)P_1 && (63)\\
&= \text{Cov}(\Delta Y_t, M|D = 1)P_1 + (\overline{\Delta Y}_1 - \overline{\Delta Y}_0)P_0 \overline{M}_1 P_1 && (64)\\
&= (\text{Cov}(\Delta Y_t, M|D = 1) + (\overline{\Delta Y}_1 - \overline{\Delta Y}_0)P_0 \overline{M}_1)P_1 && (65)
\end{aligned}
$$

We can express the denominator of (57) as:

$$\begin{aligned}
\text{Var}(D \cdot M) &= E((M \cdot D - E(M \cdot D)) \cdot (M \cdot D)) && (66)\\
&= E((M \cdot D - \overline{M}_1 P_1) \cdot (M \cdot D)) && (67)\\
&= E((M - \overline{M}_1 P_1) \cdot M|D = 1)P_1 && (68)\\
&= (E(M^2|D = 1) - \overline{M}_1^2 P_1) \cdot P_1 && (69)\\
&= (E(M^2|D = 1) - \overline{M}_1^2 P_1 - \overline{M}_1^2 P_0 + \overline{M}_1^2 P_0) \cdot P_1 && (70)\\
&= ((E(M^2|D = 1) - \overline{M}_1^2) + \overline{M}_1^2 P_0) \cdot P_1 && (71)\\
&= (\text{Var}(M|D = 1) + \overline{M}_1^2 P_0) \cdot P_1 && (72)
\end{aligned}$$

The ratio of (65) and (72) generates the following equation:

$$\frac{\text{Cov}(\Delta Y_t, D \cdot M)}{\text{Var}(D \cdot M)} = \quad = \frac{\text{Cov}(\Delta Y_t, M|D = 1) + (\overline{\Delta Y}_1 - \overline{\Delta Y}_0)P_0\overline{M}_1}{\text{Var}(M|D = 1) + \overline{M}_1^2 P_0} \tag{73}$$

If we set $\overline{M}_1 = 0$, then we have that:

$$\frac{\text{Cov}(\Delta Y_t, D \cdot M)}{\text{Var}(D \cdot M)} = \quad = \frac{\text{Cov}(\Delta Y_t, M|D = 1)}{\text{Var}(M|D = 1)} \tag{74}$$

The next part of the theorem employs the Yitzhaki's Weights (Yitzhaki 2013). Using integration by parts, it is easy to show that the covariance of any random variables Y, X such that $E(|Y|) < \infty$ and $E(|X|) = \mu_X < \infty$ and $E(Y|X)$ is differentiable, can be expressed as:

$$\text{Cov}(Y, X) = \int_{-\infty}^{\infty} \frac{\partial E(Y|X = x)}{\partial x} E(X - \mu_X|X > x)(1 - F_X(x))dx, \tag{75}$$

Moreover, we can apply (75) to express the variance of a random variable X as:

$$\text{Var}(X) \equiv \text{Cov}(X, X) = \int_{-\infty}^{\infty} E(X - \mu_X|X > x)(1 - F_X(x))dx. \tag{76}$$

Setting $\overline{M}_1 \equiv E(M|D = 1) = 0$, and applying the formula (75) to the OLS estimator in (73), we obtain:

$$\begin{aligned}
&\frac{\text{Cov}(\Delta Y_t, D \cdot M)}{\text{Var}(D \cdot M)} = \\
&= \int \frac{\partial E(\Delta Y_t|D = 1, M = m)}{\partial m} \frac{E(M|M > m, D = 1)\left(1 - F_{M|D=1}(m)\right)}{\text{Var}(M|D = 1)} dm \\
&= \int \frac{\partial E(Y_t(1) - Y_{t-1}(0)|D = 1, M = m)}{\partial m} \frac{E(M|M > m, D = 1)\left(1 - F_{M|D=1}(m)\right)}{\text{Var}(M|D = 1)} dm
\end{aligned}$$

Equation (76) and the feature that $\overline{M}_1 = 0$ assures that the weights in the equation above are always positive and integrate to one.

A.7 Proof of Theorem T.4

The first weighting scheme of the theorem simply uses the actual distribution of the data. Equation (24) is a standard result in the OLS literature. By using the full set of indicator interaction, the DiD estimator evaluates the difference of the OLS estimators if we were to regress two separate regressions, one for the control group and another for the treatment group.

Equations (25)–(26) are based on the Yitzhaki's Weights (Yitzhaki 2013), which states that the covariance of any random variables Y, X such that $E(|Y|) < \infty$ and $E(|X|) = \mu_X < \infty$ and $E(Y|X)$ is differentiable, can be expressed as:

$$\text{Cov}(Y, X) = \int_{-\infty}^{\infty} \frac{\partial E(Y|X = x)}{\partial x} \omega(x) dx,$$
$$\text{such that } \omega(x) = E(X - \mu_X | X > x)(1 - F_X(x)).$$

According to the equation above, we can express the covariances $\text{Cov}(\Delta Y_t, M|D = d); d \in \{0, 1\}$ by the following expression:

$$\text{Cov}(\Delta Y_t, M|D = d) = \int_{-\infty}^{\infty} \frac{\partial E(\Delta Y_t|D = d, M = m)}{\partial m} \omega(m) dm; \quad d \in \{0, 1\}, \tag{77}$$
$$\text{such that } \omega(m) = E(M - \overline{M}_d | M > m, D = d)(1 - F_{M|D=d}(m)),$$

where $\overline{M}_d \equiv E(M|D = d); d \in \{0, 1\}$.

The second weighting scheme sets the distribution of the moderator of the treatment and control group to the distribution of the treatment group. The DiD parameter of the regression still delivers the difference of two separate OLS regressions that evaluate the covariance between ΔY_t and M over the variance of M for each treatment group. The weighting scheme modifies the distribution of M. The first OLS parameter β_1 is associated with the treatment group ($D = 1$) and the asymptotic cumulative distribution of M is given by $F_{M|D=1}(m)$. The expected value of this OLS estimator is given by:

$$E(\beta_1) = \int_{-\infty}^{\infty} \frac{\partial E(\Delta Y_t|D = 1, M = m)}{\partial m} \omega_1(m) dm \tag{78}$$
$$\text{where } \omega_1(m) = \frac{E(M - E(M|D = 1)|M > m, D = 1)(1 - F_{M|D=1}(m))}{\text{Var}(M|D = 1)}. \tag{79}$$

The second OLS parameter β_0 is associated with the control group ($D = 0$) and the cumulative distribution of M is also given by $F_{M|D=1}(m)$. The expected value of this OLS estimator is given by:

$$E(\beta_0) = \int_{-\infty}^{\infty} \frac{\partial E(\Delta Y_t|D = 0, M = m)}{\partial m} \omega_1(m) dm, \tag{80}$$

where $\omega_1(m)$ is the same as in (79). The difference between the expected value of the OLS estimators in (86) and (80) is:

$$E(\beta_1) - E(\beta_0) = \int_{-\infty}^{\infty} \frac{\partial \left(E(\Delta Y_t | D = 1, M = m) - E(\Delta Y_t | D = 0, M = m) \right)}{\partial m} \omega_1(m) dm \tag{81}$$

$$= \int_{-\infty}^{\infty} MATT_t(m) \omega_1(m) dm, \tag{82}$$

where the second equality is due to **T.1**.

The last weighting scheme sets the conditional distribution of the moderator of the treatment and control groups to the unconditional distribution of the moderator. The DiD parameter of the regression also delivers the difference of two separate OLS regressions that evaluate the covariance between ΔY_t and M over the variance of M for each treatment group. However, the weighting scheme modifies the distribution of M. The first OLS parameter β_1^* is associated with the treatment group ($D = 1$) and the asymptotic cumulative distribution of M is given by $F_M(m)$. The expected value of this OLS estimator is given by:

$$E(\beta_1^*) = \int_{-\infty}^{\infty} \frac{\partial E(\Delta Y_t | D = 1, M = m)}{\partial m} \omega^*(m) dm \tag{83}$$

$$\text{where } \omega^*(m) = \frac{E(M - E(M|D = 1)|M > m, D = 1)(1 - F_{M|D=1}(m))}{\text{Var}(M|D = 1)}. \tag{84}$$

The second OLS parameter β_0^* is associated with the control group ($D = 0$) and the cumulative distribution of M is also given by $F_M(m)$. The expected value of this OLS estimator is given by:

$$E(\beta_0^*) = \int_{-\infty}^{\infty} \frac{\partial E(\Delta Y_t | D = 0, M = m)}{\partial m} \omega^*(m) dm, \tag{85}$$

where $\omega^*(m)$ is the same as in (84). The difference between the expected value of the two OLS estimators in (83) and (85) is:

$$E(\beta_1^*) - E(\beta_0^*) = \int_{-\infty}^{\infty} \frac{\partial \left(E(\Delta Y_t | D = 1, M = m) - E(\Delta Y_t | D = 0, M = m) \right)}{\partial m} \omega^*(m) dm \tag{86}$$

$$= \int_{-\infty}^{\infty} MATE_t(m) \omega^*(m) dm, \tag{87}$$

where the second equality is due to **T.2**.

A.8 Proof of Theorem T.5

In this proof, we adopt a concise notation to denote conditioning on $B = b$ such that $b \in \{0, 1\}$: moderator probability $P(\mathcal{M}_b) = P(M|B = b)$, the moderator expectation $\overline{M}(\mathcal{M}_b) = E(M|B = b)$, moderator variance $\text{Var}(M|\mathcal{M}_b) = \text{Var}(M|B = b)$, expectations of the outcome time-differences $\overline{\Delta Y_t}(\mathcal{M}_b) = E(\Delta Y_t|B = b)$, and covariance $\text{Cov}(\Delta Y_t, M|\mathcal{M}_b) = \text{Cov}(\Delta Y_t, M|B = b)$.

The main proof of the theorem employs the following lemmas.

Lemma L.1. If the moderator has mean zero, then the following equation holds:

$$\begin{aligned}\mathrm{Cov}(\Delta Y_T, M) &= \mathrm{Cov}(\Delta Y_t, M|\mathcal{M}_1)P(\mathcal{M}_1) + \mathrm{Cov}(\Delta Y_t, M|\mathcal{M}_0)P(\mathcal{M}_0)\\ &\quad + \overline{\Delta Y_t}(\mathcal{M}_1)\overline{M}(\mathcal{M}_1)P(\mathcal{M}_1) + \overline{\Delta Y_t}(\mathcal{M}_0)\overline{M}(\mathcal{M}_0)P(\mathcal{M}_0), \qquad (88)\\ \text{and } \mathrm{Var}(M) &= \mathrm{Var}(M|\mathcal{M}_1)P(\mathcal{M}_1) + \mathrm{Var}(M|\mathcal{M}_0)P(\mathcal{M}_0)\\ &\quad + \overline{M}(\mathcal{M}_1)^2 P(\mathcal{M}_1) + \overline{M}(\mathcal{M}_0)^2 P(\mathcal{M}_0). \qquad (89)\end{aligned}$$

Proof.

$$\begin{aligned}\mathrm{Cov}(\Delta Y, M) &= E\Big(\Delta Y \cdot \big(M \cdot B + M \cdot (1-B)\big)\Big)\\ &= E\Big(\Delta Y\big(M \cdot B\big)\Big) + E\Big(\Delta Y\big(M \cdot (1-B)\big)\Big)\\ &= E\Big(\Delta Y \cdot M|B=1\Big)P(B=1) + E\Big(\Delta Y \cdot M|B=0\Big)P(B=0)\\ &= E\Big(\Delta Y \cdot M|\mathcal{M}_1\Big)P(\mathcal{M}_1) + E\Big(\Delta Y \cdot M|\mathcal{M}_0\Big)P(\mathcal{M}_0)\\ &= \Big(E\Big(\Delta Y\big(M - \overline{M}(\mathcal{M}_1)\big|\mathcal{M}_1\Big) + \overline{\Delta Y_t}(\mathcal{M}_1)\overline{M}(\mathcal{M}_1)\Big)\Big)\, P(\mathcal{M}_1)\\ &\quad + \Big(E\Big(\Delta Y\big(M - \overline{M}(\mathcal{M}_0)\big|\mathcal{M}_0\Big) + \overline{\Delta Y_t}(\mathcal{M}_0)\overline{M}(\mathcal{M}_0)\Big)\Big)\, P(\mathcal{M}_0)\\ &= \mathrm{Cov}(\Delta Y_t, M|\mathcal{M}_1)P(\mathcal{M}_1) + \mathrm{Cov}(\Delta Y_t, M|\mathcal{M}_01)P(\mathcal{M}_0)\\ &\quad + \overline{\Delta Y_t}(\mathcal{M}_1)\overline{M}(\mathcal{M}_1)P(\mathcal{M}_1) + \overline{\Delta Y_t}(\mathcal{M}_0)\overline{M}(\mathcal{M}_0)P(\mathcal{M}_0).\end{aligned}$$

Equation (89) presents an expression for $\mathrm{Var}(M) = \mathrm{Cov}(M, M)$. Its proof is based on same rationale used for $\mathrm{Cov}(\Delta Y_t, M)$. It only requires to replace ΔY_t by M. □

Lemma L.2. The following equations hold whenever the moderator has mean zero:

$$\overline{M}(\mathcal{M}_1)^2 P(\mathcal{M}_1) + \overline{M}(\mathcal{M}_0)^2 P(\mathcal{M}_0) = \Big(\overline{M}(\mathcal{M}_1) - \overline{M}(\mathcal{M}_0)\Big)\overline{M}(\mathcal{M}_1)P(\mathcal{M}_1), \qquad (90)$$

$$\overline{\Delta Y_t}(\mathcal{M}_1)\overline{M}(\mathcal{M}_1)P(\mathcal{M}_1) + \overline{\Delta Y_t}(\mathcal{M}_0)\overline{M}(\mathcal{M}_0)P(\mathcal{M}_0) = \Big(\overline{\Delta Y_t}(\mathcal{M}_1) - \overline{\Delta Y_t}(\mathcal{M}_0)\Big)\overline{M}(\mathcal{M}_1)P(\mathcal{M}_1). \qquad (91)$$

Proof. For a zero-mean moderator, we have that:

$$\overline{M} = \overline{M}(\mathcal{M}_1)P(\mathcal{M}_1) + \overline{M}(\mathcal{M}_0)P(\mathcal{M}_0) = 0 \quad \Rightarrow \quad \overline{M}(\mathcal{M}_0)P(\mathcal{M}_0) = -\overline{M}(\mathcal{M}_1)P(\mathcal{M}_1).$$

Thus, we can rewrite equation (90) as:

$$\begin{aligned}\overline{M}(\mathcal{M}_1)^2 P(\mathcal{M}_1) + \overline{M}(\mathcal{M}_0)^2 P(\mathcal{M}_0) &= \overline{M}(\mathcal{M}_1)\overline{M}(\mathcal{M}_1)P(\mathcal{M}_1) - \overline{M}(\mathcal{M}_0)\overline{M}(\mathcal{M}_1)P(\mathcal{M}_1)\\ &= (\overline{M}(\mathcal{M}_1) - \overline{M}(\mathcal{M}_0))\overline{M}(\mathcal{M}_1)P(\mathcal{M}_1).\end{aligned}$$

The same rationale applies to equation (91). □

The following lemma applies the results in Lemmas **L.1**–**L.2**.

Lemma L.3. If the moderator has mean zero and has variance one, then the following equation hold:

$$\overline{M}(\mathcal{M}_1)P(\mathcal{M}_1) = \frac{1 - \text{Var}(M|\mathcal{M}_1)P(\mathcal{M}_1) + \text{Var}(M|\mathcal{M}_0)P(\mathcal{M}_0)}{(\overline{M}(\mathcal{M}_1) - \overline{M}(\mathcal{M}_0))}. \tag{92}$$

Proof. For a zero-mean moderator, we have that:

$$\begin{aligned}
\text{Var}(M) &= \text{Var}(M|\mathcal{M}_1)P(\mathcal{M}_1) + \text{Var}(M|\mathcal{M}_0)P(\mathcal{M}_0) + \overline{M}(\mathcal{M}_1)^2P(\mathcal{M}_1) + \overline{M}(\mathcal{M}_0)^2P(\mathcal{M}_0),\\
&= \text{Var}(M|\mathcal{M}_1)P(\mathcal{M}_1) + \text{Var}(M|\mathcal{M}_0)P(\mathcal{M}_0) + (\overline{M}(\mathcal{M}_1) - \overline{M}(\mathcal{M}_0))\overline{M}(\mathcal{M}_1)P(\mathcal{M}_1),
\end{aligned}$$

where the first equality is due to Lemma **L.1** and the second equality is due to Lemma **L.2**. If, in addition, we have that $\text{Var}(M) = 1$, then:

$$\begin{aligned}
1 &= \text{Var}(M|\mathcal{M}_1)P(\mathcal{M}_1) + \text{Var}(M|\mathcal{M}_0)P(\mathcal{M}_0) + (\overline{M}(\mathcal{M}_1) - \overline{M}(\mathcal{M}_0))\overline{M}(\mathcal{M}_1)P(\mathcal{M}_1),\\
\therefore \overline{M}(\mathcal{M}_1)P(\mathcal{M}_1) &= \frac{1 - \text{Var}(M|\mathcal{M}_1)P(\mathcal{M}_1) + \text{Var}(M|\mathcal{M}_0)P(\mathcal{M}_0)}{(\overline{M}(\mathcal{M}_1) - \overline{M}(\mathcal{M}_0))},
\end{aligned}$$

where the first expression sets $\text{Var}(M) = 1$ and second expression simply rearranges the terms. □

The lemma below applies the results in Lemmas **L.1**–**L.2** and **L.3**.

Lemma L.4. If the moderator has mean zero and has variance one, then the following equation hold:

$$\begin{aligned}
\text{Cov}(\Delta Y_T, M) &= \text{Cov}(\Delta Y_t, M|\mathcal{M}_1)P(\mathcal{M}_1) + \text{Cov}(\Delta Y_t, M|\mathcal{M}_0)P(\mathcal{M}_0) + \beta_\Delta \cdot (1 - \omega_2)\\
\text{where } \beta_\Delta &= \frac{\overline{\Delta Y_t}(\mathcal{M}_1) - \overline{\Delta Y_t}(\mathcal{M}_0)}{\overline{M}(\mathcal{M}_1) - \overline{M}(\mathcal{M}_0)},\\
\text{and } \omega_2 &= \text{Var}(M|\mathcal{M}_1)P(\mathcal{M}_1) + \text{Var}(M|\mathcal{M}_0)P(\mathcal{M}_0).
\end{aligned}$$

Proof.

$$\begin{aligned}
\text{Cov}(\Delta Y, M) &= \text{Cov}(\Delta Y_t, M|\mathcal{M}_1)P(\mathcal{M}_1) + \text{Cov}(\Delta Y_t, M|\mathcal{M}_0 1)P(\mathcal{M}_0)\\
&\quad + \overline{\Delta Y_t}(\mathcal{M}_1)\overline{M}(\mathcal{M}_1)P(\mathcal{M}_1) + \overline{\Delta Y_t}(\mathcal{M}_0)\overline{M}(\mathcal{M}_0)P(\mathcal{M}_0),\\
&= \text{Cov}(\Delta Y_t, M|\mathcal{M}_1)P(\mathcal{M}_1) + \text{Cov}(\Delta Y_t, M|\mathcal{M}_0 1)P(\mathcal{M}_0)\\
&\quad + \left(\overline{\Delta Y_t}(\mathcal{M}_1) - \overline{\Delta Y_t}(\mathcal{M}_0)\right)\overline{M}(\mathcal{M}_1)P(\mathcal{M}_1),\\
&= \text{Cov}(\Delta Y_t, M|\mathcal{M}_1)P(\mathcal{M}_1) + \text{Cov}(\Delta Y_t, M|\mathcal{M}_0 1)P(\mathcal{M}_0)\\
&\quad + \left(\overline{\Delta Y_t}(\mathcal{M}_1) - \overline{\Delta Y_t}(\mathcal{M}_0)\right) \cdot \frac{1 - \text{Var}(M|\mathcal{M}_1)P(\mathcal{M}_1) + \text{Var}(M|\mathcal{M}_0)P(\mathcal{M}_0)}{(\overline{M}(\mathcal{M}_1) - \overline{M}(\mathcal{M}_0))},\\
&= \text{Cov}(\Delta Y_t, M|\mathcal{M}_1)P(\mathcal{M}_1) + \text{Cov}(\Delta Y_t, M|\mathcal{M}_0)P(\mathcal{M}_0)\\
&\quad + \underbrace{\left(\frac{\overline{\Delta Y_t}(\mathcal{M}_1) - \overline{\Delta Y_t}(\mathcal{M}_0)}{\overline{M}(\mathcal{M}_1) - \overline{M}(\mathcal{M}_0)}\right)}_{\beta_\Delta} \cdot \underbrace{\left(1 - \text{Var}(M|\mathcal{M}_1)P(\mathcal{M}_1) + \text{Var}(M|\mathcal{M}_0)P(\mathcal{M}_0)\right)}_{1-\omega_2}.
\end{aligned}$$

where the first equality is due to Lemma **L.1**. The second equality applies Lemma **L.2** and the third equality replaces $\overline{M}(\mathcal{M}_1)P(\mathcal{M}_1)$ using Lemma **L.3**. The last equality simply rearranges the terms. □

Lemma L.5. Let the counterfactual outcome for the control units is given by

$$Y_{it}(0) = \kappa_i + \tau_t + \beta_t^0 \cdot M_i + v_{it}, \tag{93}$$

where v_{it} denotes a mean-zero exogenous error term that is statistically independent of D, B and M. Under Assumptions **A.1** and **A.2**, we have that:

$$\begin{aligned}
\frac{\text{Cov}(\Delta Y_t, M|\mathcal{M}_0)}{\text{Var}(M|\mathcal{M}_0)} &= (\beta_t^0 - \beta_{t-1}^0) && (94)\\
&= \frac{\partial E(Y_t(0) - Y_{t-1}(0)|M = m, D = 1)}{\partial m} \text{ for all } m \in \mathcal{M}. && (95)\\
&= \frac{\partial E(Y_t(0) - Y_{t-1}(0)|M = m, B = 1)}{\partial m} \text{ for all } m \in \mathcal{M}. && (96)\\
&= \frac{\partial E(Y_t(0) - Y_{t-1}(0)|M = m)}{\partial m} \text{ for all } m \in \mathcal{M}. && (97)
\end{aligned}$$

Proof. For the first equality of the lemma, recall that ΔY_{it} is defined as $\Delta Y_{it} \equiv Y_{it} - Y_{it-1}$. According to Assumption **A.2**, for any $m_0 \in \mathcal{M}_0$, we have that $Y_{it}(1, m_0) = Y_{it}(0, m_0)$ for all $i \in \mathcal{I}$. This implies that can be expressed ΔY_{it} as $\Delta Y_{it} = Y_{it}(0) - Y_{it-1}(0)$ for all $i \in \mathcal{I}$ such that $M_i \in \mathcal{M}_0$. Moreover, we have that $Y_{it}(0) = \kappa_i + \beta_t^0 \cdot M_i + v_{it}$. Thus we can rewrite ΔY_{it} as

$$\Delta Y_{it} = (\beta_t^0 - \beta_{t-1}^0) \cdot M_i + (v_{it} - v_{it-1}) \text{ for all } i \in \mathcal{I} \text{ such that } M_i \in \mathcal{M}_0. \tag{98}$$

The statistical independence relationship $(v_{it} - v_{it-1})M$ implies that:

$$\begin{aligned}
\text{Cov}(\Delta Y_t, M|\mathcal{M}_0) &= \text{Cov}((\beta_t^0 - \beta_{t-1}^0) \cdot M + (v_t - v_{t-1}), M|\mathcal{M}_0)\\
&= (\beta_t^0 - \beta_{t-1}^0)\text{Cov}(M, M|\mathcal{M}_0) + \text{Cov}((v_t - v_{t-1}), M|\mathcal{M}_0)\\
&= (\beta_t^0 - \beta_{t-1}^0)\text{Var}(M|\mathcal{M}_0) + 0\\
\Rightarrow \frac{\text{Cov}(\Delta Y_t, M|\mathcal{M}_0)}{\text{Var}(M|\mathcal{M}_0)} &= (\beta_t^0 - \beta_{t-1}^0).
\end{aligned}$$

Moreover, equation (93) enable us to express $Y_t(0) - Y_{t-1}(0)$ as:

$$Y_t(0) - Y_{t-1}(0) = (\beta_t^0 - \beta_{t-1}^0) \cdot M_i + (v_{it} - v_{it-1}) \text{ for all } i \in \mathcal{I}.$$

The statistical independence relationship $(v_{it} - v_{it-1})(D, M, B)$ holds. In particular, $(v_{it} -$

$v_{it-1})(M, D)$ implies that:

$$
\begin{aligned}
&E(Y_t(0) - Y_{t-1}(0)|D = d, M = m) = (\beta_t^0 - \beta_{t-1}^0) \cdot m, \ \forall\, (d, m) \in \{0, 1\} \times \mathcal{M} \\
\Rightarrow &\frac{\partial E(Y_t(0) - Y_{t-1}(0)|D = d, M = m)}{\partial m} = (\beta_t^0 - \beta_{t-1}^0)
\end{aligned}
$$

In addition, $(v_{it} - v_{it-1})(B, M)$ implies that:

$$
\begin{aligned}
&E(Y_t(0) - Y_{t-1}(0)|B = b, M = m) = (\beta_t^0 - \beta_{t-1}^0) \cdot m, \ \forall\, (b, m) \in \{0, 1\} \times \mathcal{M} \\
\Rightarrow &\frac{\partial E(Y_t(0) - Y_{t-1}(0)|B = b, M = m)}{\partial m} = (\beta_t^0 - \beta_{t-1}^0)
\end{aligned}
$$

Finally, $(v_{it} - v_{it-1})M$ implies that:

$$
\begin{aligned}
&E(Y_t(0) - Y_{t-1}(0)|M = m) = (\beta_t^0 - \beta_{t-1}^0) \cdot m, \ \forall\, m \in \mathcal{M} \\
\Rightarrow &\frac{\partial E(Y_t(0) - Y_{t-1}(0)|M = m)}{\partial m} = (\beta_t^0 - \beta_{t-1}^0)
\end{aligned}
$$

□

Lemma L.6. The following equation holds:

$$
\begin{aligned}
\frac{\mathrm{Cov}(\Delta Y_t, M|\mathcal{M}_1)}{\mathrm{Var}(M|\mathcal{M}_1)} &= \int \frac{\partial E(Y_t(1) - Y_{t-1}(0)|B = 1, M = m)}{\partial m} \cdot \omega(m) dm, \qquad (99) \\
\text{where } \omega(m) &= \frac{E(M - E(M|B = 1))|M > m, B = 1) \left(1 - F_{M|B=1}(m)\right)}{\mathrm{Var}(M|B = 1)}.
\end{aligned}
$$

Proof. According to Yitzhaki's Weights (Yitzhaki 2013), the covariance of any random variables Y, X such that $E(|Y|) < \infty$ and $E(|X|) = \mu_X < \infty$ and $E(Y|X)$ is differentiable, can be expressed as:

$$
\frac{\mathrm{Cov}(Y, X)}{\mathrm{Var}(X)} = \int_{-\infty}^{\infty} \frac{\partial E(Y|X = x)}{\partial x} \omega(x) dx, \qquad (100)
$$

$$
\text{such that } \omega(x) = \frac{E(X - E(X)|X > x)(1 - F_X(x))}{\mathrm{Var}(X)} \qquad (101)
$$

$$
\qquad (102)
$$

where the weighting function $\omega(x)$ is positive and integrate to one. Applying (100)–(101)

to $\frac{\text{Cov}(\Delta Y_t, M|\mathcal{M}_1)}{\text{Var}(M|\mathcal{M}_1)}$ generates the following expression:

$$\begin{aligned}
&\frac{\text{Cov}(\Delta Y_t, M|\mathcal{M}_1)}{\text{Var}(M|\mathcal{M}_1)} \\
&\quad = \frac{\text{Cov}(\Delta Y_t, M|B=1)}{\text{Var}(M|B=1)}, \\
&\quad = \int \frac{\partial E(\Delta Y_t|B=1, M=m)}{\partial m} \frac{E(M-E(M|B=1))|M>m, B=1)\left(1-F_{M|D=1}(m)\right)}{\text{Var}(M|B=1)} dm, \\
&\quad = \int \frac{\partial E(Y_t(1)-Y_{t-1}(0)|B=1, M=m)}{\partial m} \frac{E(M-E(M|B=1))|M>m, B=1)\left(1-F_{M|D=1}(m)\right)}{\text{Var}(M|B=1)} dm.
\end{aligned}$$

□

The following Lemma applies the results in Lemmas L.5 and L.6:

Lemma L.7. Let the parameter β_{MATT} be defined as:

$$\beta_{MATT} \equiv \int \frac{\partial E(Y_t(1)-Y_t(0)|B=1, M=m)}{\partial m} \omega(m) dm, \tag{103}$$
$$\text{where } \omega(m) = \frac{E(M-E(M|B=1)|M>m, B=1)\left(1-F_{M|B=1}(m)\right)}{\text{Var}(M|B=1)}.$$

Under Assumptions A.1–A.2, if the counterfactual outcome for the control units is given by (31) in T.5, then the following equation holds:

$$\beta_{MATT} = \frac{\text{Cov}(\Delta Y_t, M|\mathcal{M}_1)}{\text{Var}(M|\mathcal{M}_1)} - \frac{\text{Cov}(\Delta Y_t, M|\mathcal{M}_0)}{\text{Var}(M|\mathcal{M}_0)} \tag{104}$$

Proof. Lemma L.5 shows that $\frac{\partial E(Y_t(0)-Y_{t-1}(0)|M=m, D=1)}{\partial m}$ is constant and given by:

$$\frac{\partial E(Y_t(0)-Y_{t-1}(0)|M=m, D=1)}{\partial m} = (\beta_t^0 - \beta_{t-1}^0) \text{ for all } m \in \mathcal{M} = \frac{\partial E(Y_t(0)-Y_{t-1}(0)|M=m, B=1)}{\partial m}.$$

Lemma L.6 states that

$$\frac{\text{Cov}(\Delta Y_t, M|\mathcal{M}_1)}{\text{Var}(M|\mathcal{M}_1)} = \int \frac{\partial E(Y_t(1)-Y_{t-1}(0)|B=1, M=m)}{\partial m} \cdot \omega(m) dm,$$

where the weights $\omega(m)$ are positive and integrate to one, $\int \omega(m) dm = 1$. Thus, we have that:

$$\begin{aligned}
(\beta_t^0 - \beta_{t-1}^0) &= \int (\beta_t^0 - \beta_{t-1}^0) \omega(m) dm \\
&= \int \frac{\partial E(Y_t(1)-Y_{t-1}(0)|B=1, M=m)}{\partial m} \omega(m) dm.
\end{aligned}$$

By combining these results, we have that:

$$
\begin{aligned}
&\frac{\mathrm{Cov}(\Delta Y_t, M|\mathcal{M}_1)}{\mathrm{Var}(M|\mathcal{M}_1)} - \frac{\mathrm{Cov}(\Delta Y_t, M|\mathcal{M}_0)}{\mathrm{Var}(M|\mathcal{M}_0)} \\
&\quad = \int \frac{\partial E(Y_t(1) - Y_{t-1}(0)|B=1, M=m)}{\partial m} \cdot \omega(m)dm - \int \frac{\partial E(Y_t(0) - Y_{t-1}(0)|B=1, M=m)}{\partial m}\omega(m)dm \\
&\quad = \int \frac{\partial E(Y_t(1) - Y_{t-1}(0)|B=1, M=m) - E(Y_t(0) - Y_{t-1}(0)|B=1, M=m)}{\partial m} \cdot \omega(m)dm \\
&\quad = \int \frac{\partial E(Y_t(1) - Y_{t-1}(0)|B=1, M=m)}{\partial m} \cdot \omega(m)dm,
\end{aligned}
$$

□

We are now equipped to prove the theorem:

$$
\begin{aligned}
\mathrm{Cov}(\Delta Y_T, M) &= \mathrm{Cov}(\Delta Y_t, M|\mathcal{M}_1)P(\mathcal{M}_1) + \mathrm{Cov}(\Delta Y_t, M|\mathcal{M}_0)P(\mathcal{M}_0) + \beta_\Delta \cdot (1-\omega_2) \\
&= \frac{\mathrm{Cov}(\Delta Y_t, M|\mathcal{M}_1)}{\mathrm{Var}(M|\mathcal{M}_1)}\mathrm{Var}(M|\mathcal{M}_1)P(\mathcal{M}_1) + \mathrm{Cov}(\Delta Y_t, M|\mathcal{M}_0)P(\mathcal{M}_0) + \beta_\Delta \cdot (1-\omega_2) \\
&= \left(\frac{\mathrm{Cov}(\Delta Y_t, M|\mathcal{M}_1)}{\mathrm{Var}(M|\mathcal{M}_1)} - \frac{\mathrm{Cov}(\Delta Y_t, M|\mathcal{M}_0)}{\mathrm{Var}(M|\mathcal{M}_0)}\right)\mathrm{Var}(M|\mathcal{M}_1)P(\mathcal{M}_1) \\
&\qquad + \frac{\mathrm{Cov}(\Delta Y_t, M|\mathcal{M}_0)}{\mathrm{Var}(M|\mathcal{M}_0)}\mathrm{Var}(M|\mathcal{M}_1)P(\mathcal{M}_1) + \mathrm{Cov}(\Delta Y_t, M|\mathcal{M}_0)P(\mathcal{M}_0) + \beta_\Delta \cdot (1-\omega_2) \\
&= \left(\frac{\mathrm{Cov}(\Delta Y_t, M|\mathcal{M}_1)}{\mathrm{Var}(M|\mathcal{M}_1)} - \frac{\mathrm{Cov}(\Delta Y_t, M|\mathcal{M}_0)}{\mathrm{Var}(M|\mathcal{M}_0)}\right)\mathrm{Var}(M|\mathcal{M}_1)P(\mathcal{M}_1) \\
&\qquad + \frac{\mathrm{Cov}(\Delta Y_t, M|\mathcal{M}_0)}{\mathrm{Var}(M|\mathcal{M}_0)}\left(\mathrm{Var}(M|\mathcal{M}_1)P(\mathcal{M}_1) + P(\mathcal{M}_0)\mathrm{Var}(M|\mathcal{M}_0)\right) + \beta_\Delta \cdot (1-\omega_2) \\
&= \left(\frac{\mathrm{Cov}(\Delta Y_t, M|\mathcal{M}_1)}{\mathrm{Var}(M|\mathcal{M}_1)} - \frac{\mathrm{Cov}(\Delta Y_t, M|\mathcal{M}_0)}{\mathrm{Var}(M|\mathcal{M}_0)}\right)\mathrm{Var}(M|\mathcal{M}_1)P(\mathcal{M}_1) \\
&\qquad + \frac{\mathrm{Cov}(\Delta Y_t, M|\mathcal{M}_0)}{\mathrm{Var}(M|\mathcal{M}_0)}\omega_2 + \beta_\Delta \cdot (1-\omega_2) \\
&= \beta_{MATT} \cdot \mathrm{Var}(M|\mathcal{M}_1)P(\mathcal{M}_1) + (\beta_t^0 - \beta_{t-1}^0)\omega_2 + \beta_\Delta \cdot (1-\omega_2),
\end{aligned}
$$

where the first equality is due to Lemma **L.4**. The second, third, fourth and fifth equalities only perform simple algebraic manipulations. The last equation is due to Lemma **L.7**.

In the case that $P(B=1)=1$, then $B=D$ and $P(D=1)=1$ also holds. Moreover, the covariance $\mathrm{Cov}(\Delta Y_T, M)$ can be expressed as:

$$
\mathrm{Cov}(\Delta Y_T, M) = \mathrm{Cov}(\Delta Y_t, M|B=1)P(B=1) = \mathrm{Cov}(\Delta Y_t, M|D=1)P(D=1).
$$

Recall that M is standardized and thereby $\mathrm{Var}(M) = \mathrm{Var}(M|D=1) = \mathrm{Var}(M|B=1) = 1$.

Thus we have that:

$$\begin{aligned}
\text{Cov}(\Delta Y_T, M) &= \frac{\text{Cov}(\Delta Y_t, M|B=1)}{\text{Var}(M|B=1)}, \\
&= \frac{\text{Cov}(\Delta Y_t, M|\mathcal{M}_1)}{\text{Var}(M|\mathcal{M}_1)}, \\
&= \int \frac{\partial E(Y_t(1) - Y_{t-1}(0)|B=1, M=m)}{\partial m} \cdot \omega(m) dm, \\
\text{where } \omega(m) &= \frac{E(M - E(M|B=1))|M > m, B=1)\left(1 - F_{M|B=1}(m)\right)}{\text{Var}(M|B=1)},
\end{aligned}$$

where the second equality is due to Lemma L.6. Suppose that is no time trend, that is, $\beta_t^0 = \beta_{t-1}^0$. According to LemmaL.5, we have that:

$$\begin{aligned}
&E(Y_t(0) - Y_{t-1}(0)|B=b, M=m) = (\beta_t^0 - \beta_{t-1}^0) \cdot m, \ \forall\, (b,m) \in \{0,1\} \times \mathcal{M} \\
\Rightarrow\ &\frac{\partial E(Y_t(0) - Y_{t-1}(0)|B=b, M=m)}{\partial m} = (\beta_t^0 - \beta_{t-1}^0) = 0
\end{aligned}$$

Thus:

$$\begin{aligned}
\text{Cov}(\Delta Y_T, M) &= \int \frac{\partial E(Y_t(1) - Y_{t-1}(0)|B=1, M=m) - 0}{\partial m} \cdot \omega(m) dm, \\
&= \int \frac{\partial E(Y_t(1) - Y_{t-1}(0)|B=1, M=m) - \frac{\partial E(Y_t(0) - Y_{t-1}(0)|B=b, M=m)}{\partial m}}{\partial m} \cdot \omega(m) dm, \\
&= \int \frac{\partial E(Y_t(1) - Y_t(0)|B=1, M=m)}{\partial m} \cdot \omega(m) dm, \\
&= \int \frac{\partial E(Y_t(1) - Y_t(0)|D=1, M=m)}{\partial m} \cdot \omega_D(m) dm, \\
\text{where } \omega_D(m) &= \frac{E(M - E(M|D=1))|M > m, D=1)\left(1 - F_{M|D=1}(m)\right)}{\text{Var}(M|D=1)}, \\
&= \int MATT(m) \cdot \frac{E(M - E(M|D=1))|M > m, D=1)\left(1 - F_{M|D=1}(m)\right)}{\text{Var}(M|D=1)} dm,
\end{aligned}$$

where the fourth equality uses tha fact that $D = B$ when $P(B=1) = 1$.

Appendix B Additional Tables

TABLE A1: IMPACT OF PNTR ON MENTAL HEALTH INDEX

	(1)	(2)	(3)
Post × NTR gap	0.061*	0.071**	0.067**
	(0.033)	(0.033)	(0.033)
Observations	3549	3549	3549
Outcome mean	0.00	0.00	0.00
Birth prefecture fixed effects	Yes	Yes	Yes
Birth year fixed effects	Yes	Yes	Yes
Survey year fixed effects	Yes	Yes	Yes
Birth year fixed effects			
× Other trade policies	Yes	Yes	Yes
× Initial prefecture characteristics		Yes	Yes
Individual characteristics			Yes

Notes: Data are from the 2016–2018 CFPS. This table reports results of the difference-in-difference regressions of adolescent mental health index on interaction of the prefecture-level NTR gap and a post-PNTR indicator. Regressions in column 1 control for birth year fixed effects, prefecture of birth fixed effects, survey year fixed effects, and the birth year fixed effects interacted with other trade policies including China's output tariffs, input tariffs, the NTR rates, export licensing, MFA quotas, and contract intensity. Regressions in column 2 further control for the birth year fixed effects interacted with initial prefecture characteristics including GDP per capita, share of employment in manufacturing, and distance to the nearest port. Regressions in column 3 further control for individual characteristics including adolescent age fixed effects, adolescent gender, and indicator variables for whether the mother and father completed middle school. Standard errors are clustered at the prefecture of birth level. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

Table A2: Impact of PNTR on Adolescent Physical Health, Cognition and School Dropout

	(1)	(2)	(3)
Panel A: Physical health index			
Post × NTR gap	0.020	0.024	0.030
	(0.037)	(0.039)	(0.038)
Observations	3547	3547	3547
Outcome mean	0.00	0.00	0.00
Panel B: Cognitive function index			
Post × NTR gap	-0.019	0.006	0.012
	(0.035)	(0.035)	(0.033)
Observations	2984	2984	2984
Outcome mean	0.00	0.00	0.00
Panel C: School dropout			
Post × NTR gap	-0.012	-0.009	-0.010
	(0.011)	(0.011)	(0.011)
Observations	3549	3549	3549
Outcome mean	0.10	0.10	0.10
Birth prefecture fixed effects	Yes	Yes	Yes
Birth year fixed effects	Yes	Yes	Yes
Survey year fixed effects	Yes	Yes	Yes
Birth year fixed effects			
× Other trade policies	Yes	Yes	Yes
× Initial prefecture characteristics		Yes	Yes
Individual characteristics			Yes

Notes: Data are from the 2016–2018 CFPS. This table reports results of the difference-in-difference regressions of child physical health and cognitive function on interaction of the prefecture-level NTR gap and a post-PNTR indicator. Regressions in column 1 control for birth year fixed effects, prefecture of birth fixed effects, survey year fixed effects, and the birth year fixed effects interacted with other trade policies including China's output tariffs, input tariffs, the NTR rates, export licensing, MFA quotas, and contract intensity. Regressions in column 2 further control for the birth year fixed effects interacted with initial prefecture characteristics including GDP per capita, share of employment in manufacturing, and distance to the nearest port. Regressions in column 3 further control for individual characteristics including adolescent age fixed effects, adolescent gender, and indicator variables for whether the mother and father completed middle school. Standard errors are clustered at the prefecture of birth level. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

Table A3: Robustness Checks: Alternative Measures of NTR Gap

	Mild depression (1)	Severe depression (2)
Panel A: NTR gap measured by excluding industries with the highest value of NTR gap		
Post × NTR gap	0.011	-0.025***
	(0.012)	(0.009)
Observations	3549	3549
Outcome mean	0.16	0.06
Panel B: NTR gap measured by excluding industries with the lowest value of NTR gap		
Post × NTR gap	0.016	-0.032***
	(0.015)	(0.012)
Observations	3549	3549
Outcome mean	0.16	0.06
Panel C: NTR gap winsorized at the 5/95 percentiles		
Post × NTR gap	0.015	-0.030***
	(0.015)	(0.011)
Observations	3549	3549
Outcome mean	0.16	0.06
Panel D: NTR gap measured by excluding nontradable industries		
Post × NTR gap	0.011	-0.024**
	(0.013)	(0.010)
Observations	3549	3549
Outcome mean	0.16	0.06
Birth prefecture fixed effects	Yes	Yes
Birth year fixed effects	Yes	Yes
Survey year fixed effects	Yes	Yes
Birth year fixed effects		
× Other trade policies	Yes	Yes
× Initial prefecture characteristics	Yes	Yes
Individual characteristics	Yes	Yes

Notes: Data are from the 2016–2018 CFPS. This table reports results of the difference-in-difference regressions of adolescent mental health on interaction of the prefecture-level NTR gap and a post-PNTR indicator. All regressions control for birth year fixed effects, prefecture of birth fixed effects, survey year fixed effects, the birth year fixed effects interacted with other trade policies including China's output tariffs, input tariffs, the NTR rates, export licensing, MFA quotas, and contract intensity, the birth year fixed effects interacted with initial prefecture characteristics including GDP per capita, share of employment in manufacturing, and distance to the nearest port, and individual characteristics including adolescent age fixed effects, adolescent gender, and indicator variables for whether the mother and father completed middle school. Knitwear industry has the highest NTR gaps and is excluded in Panel A. Water resources management industry, coal mining and washing industry, mineral mining and processing industry, and coking industry have the lowest NTR gaps and are excluded in Panel B. Standard errors are clustered at the prefecture of birth level. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

Table A4: Robustness Checks: Alternative Specifications and Alternative Samples

	Mild depression (1)	Severe depression (2)
Panel A: Regression weighted by 1990 prefecture population		
Post × NTR gap	-0.005	-0.031***
	(0.016)	(0.012)
Observations	3549	3549
Outcome mean	0.16	0.06
Panel B: Controlling for prefecture of birth-year linear trends		
Post × NTR gap	0.053*	-0.041**
	(0.027)	(0.021)
Observations	3549	3549
Outcome mean	0.16	0.06
Panel C: Using cohorts born between 1997 and 2006		
Post × NTR gap	0.010	-0.016**
	(0.011)	(0.008)
Observations	5978	5978
Outcome mean	0.15	0.06
Birth prefecture fixed effects	Yes	Yes
Birth year fixed effects	Yes	Yes
Survey year fixed effects	Yes	Yes
Individual characteristics	Yes	Yes

Notes: Data are from the 2016–2018 CFPS. This table reports results of the difference-in-difference regressions of adolescent mental health on interaction of the prefecture-level NTR gap and a post-PNTR indicator. All regressions control for birth year fixed effects, prefecture of birth fixed effects, survey year fixed effects, the birth year fixed effects interacted with other trade policies including China's output tariffs, input tariffs, the NTR rates, export licensing, MFA quotas, and contract intensity, the birth year fixed effects interacted with initial prefecture characteristics including GDP per capita, share of employment in manufacturing, and distance to the nearest port, and individual characteristics including adolescent age fixed effects, adolescent gender, and indicator variables for whether the mother and father completed middle school. Regressions in Panel A are weighted by 1990 prefecture population. Regressions in Panel B further control for prefecture- year of birth linear trend. Regressions in Panel C use children born between 1997 and 2006. Standard errors are clustered at the prefecture of birth level. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

Table A5: Heterogeneous Effects of PNTR on Adolescent Mental Health Outcomes

	Mild depression (1)	Severe depression (2)
Panel A: Interact with "Female"		
Post × NTR gap	0.010	-0.025***
	(0.012)	(0.009)
Post × NTR gap × interaction	0.002	0.001
	(0.002)	(0.001)
Observations	3549	3549
Outcome mean	0.16	0.06
Panel B: Interact with "Mother completed middle school"		
Post × NTR gap	0.011	-0.025***
	(0.012)	(0.009)
Post × NTR gap × interaction	-0.003	0.002
	(0.003)	(0.001)
Observations	3549	3549
Outcome mean	0.16	0.06
Panel C: Interact with "Father completed middle school"		
Post × NTR gap	0.010	-0.025***
	(0.012)	(0.009)
Post × NTR gap × interaction	0.001	-0.001
	(0.003)	(0.001)
Observations	3549	3549
Outcome mean	0.16	0.06
Panel D: Interact with "Parental absence for at least one week during ages 0-3"		
Post × NTR gap	0.005	-0.025***
	(0.012)	(0.009)
Post × NTR gap × interaction	-0.000	0.000
	(0.003)	(0.002)
Observations	3375	3375
Outcome mean	0.16	0.06
Panel E: Interact with "Above median initial share of rural population"		
Post × NTR gap	0.010	-0.020**
	(0.013)	(0.010)
Post × NTR gap × interaction	0.000	-0.003
	(0.003)	(0.002)
Observations	3549	3549
Outcome mean	0.16	0.06
Birth prefecture fixed effects	Yes	Yes
Birth year fixed effects	Yes	Yes
Survey year fixed effects	Yes	Yes
Birth year fixed effects		
× Other trade policies	Yes	Yes
× Initial prefecture characteristics	Yes	Yes
Individual characteristics	Yes	Yes

Notes: Data are from the 2016–2018 CFPS. This table reports results of the difference-in-difference regressions of adolescent mental health on interaction of the prefecture-level NTR gap and a post-PNTR indicator, and a triple interaction of that term with a female indicator in Panel A, with an indicator for whether mother completed middle school in Panel B, with an indicator for whether father completed middle school in Panel C, and with an indicator of parental absence for at least one week during ages 0-3 in Panel D. All regressions control for birth year fixed effects, prefecture of birth fixed effects, survey year fixed effects, and the birth year fixed effects interacted with other trade policies including China's output tariffs, input tariffs, the NTR rates, export licensing, MFA quotas, and contract intensity, the birth year fixed effects interacted with initial prefecture characteristics including GDP per capita, share of employment in manufacturing, and distance to the nearest port, and individual characteristics including adolescent age fixed effects, adolescent gender, and indicator variables for whether the mother and father completed middle school. Standard errors are clustered at the prefecture of birth level. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

Table A6: Impact of PNTR on Migration

	(1)	(2)	(3)
Panel A: Migrated since birth			
Post × NTR gap	-0.004	-0.005	-0.005
	(0.003)	(0.003)	(0.004)
Observations	3388	3388	3388
Outcome mean	0.01	0.01	0.01
Panel B: Migrated since age 3			
Post × NTR gap	-0.003	-0.004	-0.003
	(0.003)	(0.003)	(0.003)
Observations	3283	3283	3283
Outcome mean	0.01	0.01	0.01
Panel C: Migrated from rural to urban region			
Post × NTR gap	-0.008	-0.005	-0.005
	(0.006)	(0.008)	(0.008)
Observations	3369	3369	3369
Outcome mean	0.04	0.04	0.04
Birth prefecture fixed effects	Yes	Yes	Yes
Birth year fixed effects	Yes	Yes	Yes
Survey year fixed effects	Yes	Yes	Yes
Birth year fixed effects			
× Other trade policies	Yes	Yes	Yes
× Initial prefecture characteristics		Yes	Yes
Individual characteristics			Yes

Notes: Data are from the 2016–2018 CFPS. This table reports results of the difference-in-difference regressions of migration indicators on interaction of the prefecture-level NTR gap and a post-PNTR indicator. Regressions in column 1 control for birth year fixed effects, prefecture of birth fixed effects, survey year fixed effects, and the birth year fixed effects interacted with other trade policies including China's output tariffs, input tariffs, the NTR rates, export licensing, MFA quotas, and contract intensity. Regressions in column 2 further control for the birth year fixed effects interacted with initial prefecture characteristics including GDP per capita, share of employment in manufacturing, and distance to the nearest port. Regressions in column 3 further control for individual characteristics including adolescent age fixed effects, adolescent gender, and indicator variables for whether the mother and father completed middle school. Standard errors are clustered at the prefecture of birth level. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

Table A7: Impact of PNTR on Parental Absence

	(1)	(2)	(3)
Panel A: Parents have been absent for at least one week during ages 0-3			
Post × NTR gap	-0.012	-0.006	-0.007
	(0.016)	(0.018)	(0.018)
Observations	3375	3375	3375
Outcome mean	0.18	0.18	0.18
Birth prefecture fixed effects	Yes	Yes	Yes
Birth year fixed effects	Yes	Yes	Yes
Survey year fixed effects	Yes	Yes	Yes
Birth year fixed effects			
× Other trade policies	Yes	Yes	Yes
× Initial prefecture characteristics		Yes	Yes
Individual characteristics			Yes
Panel B: Parent was not living in the household and seeking employment elsewhere			
Panel B1: Mother			
Post × NTR gap	-0.002	0.001	0.001
	(0.004)	(0.004)	(0.004)
Observations	2135	2135	2135
Outcome mean	0.01	0.01	0.01
Panel B2: Father			
Post × NTR gap	0.002	0.004	0.003
	(0.004)	(0.005)	(0.005)
Observations	2133	2133	2133
Outcome mean	0.02	0.02	0.02
Prefecture fixed effects	Yes	Yes	Yes
Birth year fixed effects	Yes	Yes	Yes
Survey year fixed effects	Yes	Yes	Yes
Birth year fixed effects			
× Other trade policies	Yes	Yes	Yes
× Initial prefecture characteristics		Yes	Yes
Individual characteristics			Yes

Notes: Data in Panel A are from the 2016–2018 CFPS and data in Panel B are from the 2000-2015 CHNS. Regressions in column 1 control for birth year fixed effects, birth prefecture fixed effects (prefecture fixed effects in Panel B), survey year fixed effects, and the birth year fixed effects interacted with other trade policies including China's output tariffs, input tariffs, the NTR rates, export licensing, MFA quotas, and contract intensity. Regressions in column 2 further control for the birth year fixed effects interacted with initial prefecture characteristics including GDP per capita, share of employment in manufacturing, and distance to the nearest port. Regressions in column 3 further control for individual characteristics including adolescent age fixed effects, adolescent gender, and indicator variables for whether the mother and father completed middle school. Standard errors are clustered at the prefecture of birth level in Panel A and are clustered at the prefecture level in Panel B. ***, **, and * denote significance at the 1, 5, and 10 percent levels.

Table A8: Impact of PNTR on Fertility Outcomes

	Births per 1,000 women (1)	Number of children (2)
Post × NTR gap	0.261	89.945
	(0.462)	(677.960)
Observations	1655	1655
Outcome mean	45.52	7644.84
Prefecture fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Year fixed effects		
× Other trade policies	Yes	Yes
× Initial prefecture characteristics	Yes	Yes

Notes: Data are from the 1990, 2000, 2005, 2010, and 2015 population census in China. This table reports results of the difference-in-difference regressions of fertility on interaction of the prefecture-level NTR gap and a post-PNTR indicator. All regressions control for prefecture fixed effects, year fixed effects, the year fixed effects interacted with other trade policies including China's output tariffs, input tariffs, the NTR rates, export licensing, MFA quotas, and contract intensity, and the survey year fixed effects interacted with initial prefecture characteristics including GDP per capita, share of employment in manufacturing, and distance to the nearest port. All regressions are weighted by 1990 prefecture population. Standard errors are clustered at the prefecture level. ***, **, and * denote significance at the 1, 5, and 10 percent levels.